THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 2, February 1981
*Printed in U.S.A.*

# The Quality Measurement Plan (QMP)

By BRUCE HOADLEY

(Manuscript received August 5, 1980)

*This paper describes the Quality Measurement Plan (QMP), a recently implemented system for reporting the quality assurance audit results to Bell System management. QMP replaces the T-rate system, which evolved from the pioneering statistical work of Shewhart and Dodge during the 1920's and 1930's at Bell Laboratories. Box and whisker plots are used for graphically displaying confidence intervals for the quality of the current production. The confidence interval is computed from both current and past data and is derived from a new Bayesian approach to the empirical Bayes problem for Poisson observations. Here we discuss the rationale, mathematical derivations, dynamics, operating characteristics, and many comparative examples. We show that QMP reduces statistical errors relative to the earlier T-rate system.*

## I. INTRODUCTION

### *1.1 Quality assurance*

The responsibility of the Bell Laboratories Quality Assurance Center (QAC) is "to ensure that the communications products designed by Bell Laboratories and bought by Bell System operating companies from Western Electric Company, Incorporated will meet quality standards and will perform as the designers intended."[1] This obviates the need for each operating company to carry out its own acceptance inspection.

To meet this responsibility, the QAC works with its Western Electric (WE) agents, the Quality Assurance Directorate (QAD),[2] and Purchased Products Inspection (PPI) organizations. However, as stated in Ref. 1, "The primary responsibility for quality lies with the line organizations: Bell Laboratories for the quality of design and Western Electric for the quality of manufacture, installation, and repair." The quality assurance organizations conduct independent activities to assure quality to the operating companies.

### 1.2 Quality assurance audit

The quality assurance organizations have two major activities. The first is to conduct quality audits where products change hands, either within WE or between WE and the operating companies. Examples are manufacturing, installation, and repair audits. The second concerns a collection of field quality monitoring activities. Examples are the Product Performance Surveys. These are designed sample surveys of reported field troubles.

An audit is a highly structured system of inspections done on a sampling basis. The ingredients of an audit are: (*i*) sampling method, (*ii*) scope of inspection, (*iii*) quality standards, (*iv*) nonconformance procedures, (*v*) defect assessment practices, (*vi*) quality rating method, and (*vii*) report formats.

The sampling method along with the scope of inspection determines what tests will be performed on what units of product or attributes of product. The statistics and economics of sampling, the engineering requirements, and the field effect of defects play the central roles in determining the sampling and the scope of inspection.

The quality standards are numerical values expressed in defects, defectives or demerits per unit. They are set by the QAC in consultation with the QAD. The standards are target values, reflecting a tradeoff between first cost and maintenance costs.

The nonconformance procedures are rules for detecting and disposing of audited lots that are excessively defective with respect to a particular set of engineering requirements.

The defect assessment practices are a set of transformations that map defects found into defects assessed for quality rating purposes. A terminal strip may have all ten connections off by one position, but, the consequences of these ten defects found are much less than ten independent occurrences of this defect. Therefore, less than ten defects are assessed.

The quality rating method and report formats determine how the results of the audit are presented to Bell System management. For example, a product is reported as "Below Normal," when it fails a statistical test of the hypothesis that the quality standard is being met.

### 1.3 The quality measurement plan (QMP)

The statistical foundations of the audit ingredients were developed by Shewhart, Dodge, and others, starting in the 1920's and continuing through to the middle 1950's. This work was documented in the literature in Refs. 3 to 6.

In recent years, research has been carried out to evaluate the application of modern statistical theories to the audit ingredients. An important idea is summarized in an article by Efron and Morris[7] which explains a paradox discovered by Stein.[8] When you have samples from similar populations, the individual sample characteristics are not the best estimates of the individual population characteristics. Total error is reduced by shrinking the individual sample characteristics part way towards the grand mean over all samples. Efron and Morris used baseball batting averages to illustrate the point. But the problem of estimating percent defective in quality assurance is the same problem. And you are always concerned with similar populations—for example, the population of design-line telephones produced for each of several months.

This idea was originally explored in Ref. 9. The idea has now evolved into the Quality Measurement Plan (QMP). QMP is the recently implemented system for conducting three of the audit ingredients: defect assessment, quality rating, and quality reporting.

As a quick introduction to QMP, consider Fig. 1. This is a comparison of the QMP reporting format (Fig. 1a) with the old $T$-rate reporting format (Fig. 1b). Each year is divided into eight periods. On the bottom, the $T$-rate is plotted for each period and it measures the difference between the observed and standard defect rates in units of sampling standard deviation (given standard quality). The idea is that if the $T$-rate is, e.g., less than negative two, then the hypothesis of standard quality is rejected. Section II considers the exact rules for exception reporting under the $T$-rate system.

Under QMP, a box and whisker chart is plotted each period. The box chart is a graphical representation of the posterior distribution of current population quality on an index scale. The index value one is the standard on the index scale and the value two means twice as many defects as expected under the quality standard. The posterior probability that the population index is larger than the top whisker is 0.99. The top of the box, the bottom of the box, and the bottom whisker correspond to the probabilities 0.95, 0.05, and 0.01, respectively.

The heavy "dot" is a Bayes estimate of the long run *process average*; the "cross" is the observed value in the current sample; and the "dash" near the middle of the box is the posterior mean of the current population index and is called the *Best Measure* of current quality. The process averages, "dots," are joined to show trends.

Although the $T$-rate chart and the QMP chart often convey similar messages, there are differences. The QMP chart provides a measure of quality; the $T$-rate chart does not. For example, in 7806 (Period 6 of 1978) both charts imply that the quality is substandard, but the QMP chart also implies that the population index is somewhere between one and two.

QMP and the $T$-rate use the past data in very different ways. QMP

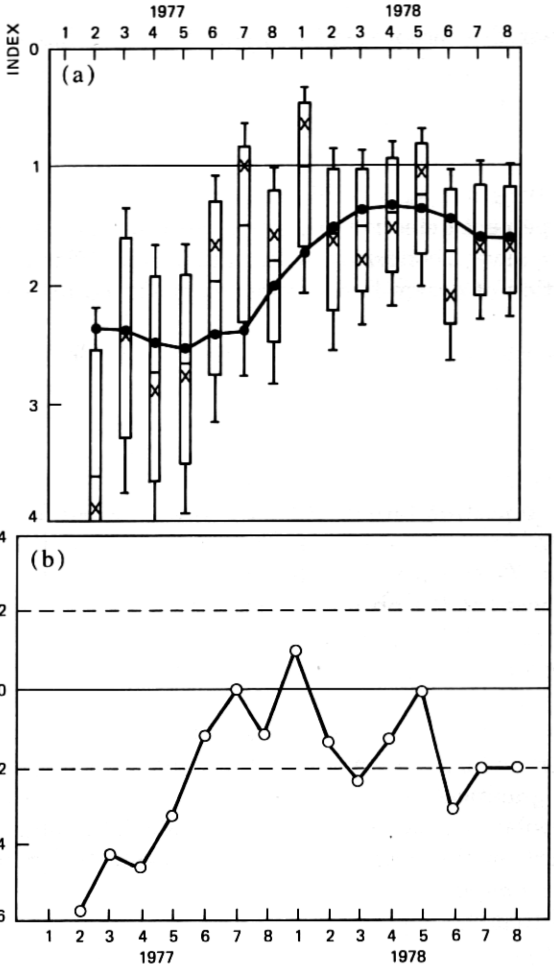


Fig. 1—QMP versus the *T*-rate. The box and whisker plot in (a) is the QMP replacement of the *T*-rate. One is standard on the index scale; two is a defect rate of twice standard. The box and whisker are 90 and 98 percent confidence intervals for production during the period; the "crosses" are the indices in the samples; the "dots" are process averages; and the "dashes" in the middle of the boxes are Best Measures of current quality derived from empirical Bayes theory. (b) is a time series of *T*-rates. Each point measures the difference between observed quality and expected quality on a standard deviation scale. Notice that the sixth period of 1977 and the fourth period of 1978 are the same in the *T*-rate chart but quite different in the QMP chart.

uses the past sample indices, but makes an inference about current quality. The *T*-rate system uses runs criteria based on attributes of the *T*-rate, such as "less than zero," and can make an inference about past quality. In Fig. 2, 7707, the *T*-rate signals an exception, because six *T*-rates in a row are less than zero, indicating that quality has not been standard for all six periods. But for QMP, the standard is well within the box, indicating normal current quality. The different treatment of past data is also illustrated in Fig. 1. Comparing 7706 with 7804 reveals very similar *T*-rates, but QMP box charts with different messages.

The *T*-rate system is based on the assumption that the total number of defects in a rating period has a normal distribution. QMP is based on the Poisson distribution. This difference is important for small audits, as shown in Section VII.

QMP was on trial for two years and was applied to 20,000 sets of audit data. The relatively simple QMP algorithm published in Ref. 9

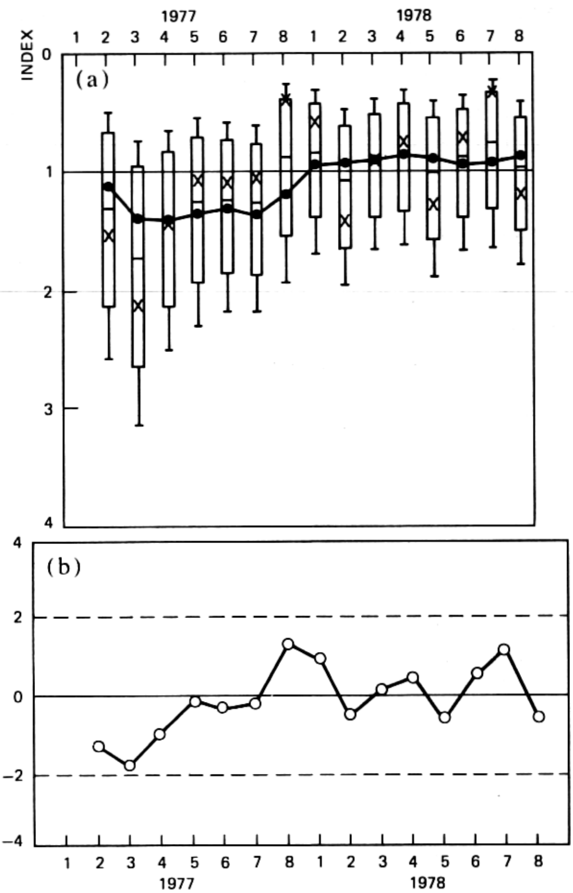


Fig. 2—A weak *T*-rate exception. The seventh period of 1977 was reported as a quality exception because six *T*-rates in a row were less than zero. For QMP, it would have been reported as normal. This is because QMP provides a statistical inference about current production only, even though past data is used.

was used originally. This simple algorithm worked for most data sets, but not all (e.g., zero defects in every period). The relatively complex algorithm discussed in Section IV is the result of a lengthy fine-tuning process, designed to make the algorithm work for every case. This is why the full power of Bayes theorem with empirically based prior distributions had to be used.

### 1.4 Relationship to the empirical Bayes approach

Note that in the QMP box chart, the Best Measure always lies between the estimated process average and the current sample index. The Best Measure is a shrinkage of the sample index towards the estimated process average. In 7706 of Fig. 1a, the shrinkage is away from standard; but, in 7804, it is towards the standard.

The Best Measure is related to the class of estimators described by Efron and Morris.[10] In the cited reference, they provide a foundation for Stein's paradox with an empirical Bayes approach. In Ref. 7, they used baseball data to illustrate Stein's paradox. There is a clear analogy between percent defective in a quality assurance application and a baseball batting average. The data in Ref. 7 was for many players at a given point in time. The QMP algorithm works with the data for one product over time. So a better baseball analogy would be one player over time.

Table I contains batting average data for Thurman Munson from 1970 through 1978. This data was collected and analyzed by S. G. Crawford and is displayed graphically in Fig. 3. The "crosses" are Munson's batting averages reported on the last Sunday of April for each year. The "boxes" are Munson's batting averages at the end of the season. The dashed line is the average of the "crosses."

The early season averages are analogous to the audit data. The averages are the results from small samples of the populations. The populations are the finite populations of "at bats" for each season. In

Table I—Batting average data for Thurman Munson

| | Reported* on Last Sunday in April | | | End of Season | | | QMP Estimate of Batting Average |
|---|---|---|---|---|---|---|---|
| Year | At Bats | Hits | Batting Average | At Bats | Hits | Batting Average | |
| 1970 | 44 | 7 | 0.159 | 453 | 137 | 0.302 | 0.165 |
| 1971 | 37 | 6 | 0.162 | 451 | 113 | 0.251 | 0.168 |
| 1972 | 31 | 11 | 0.355 | 511 | 143 | 0.280 | 0.288 |
| 1973 | 56 | 20 | 0.357 | 519 | 156 | 0.301 | 0.315 |
| 1974 | 72 | 16 | 0.222 | 517 | 135 | 0.261 | 0.233 |
| 1975 | 48 | 18 | 0.375 | 597 | 190 | 0.318 | 0.323 |
| 1976 | 40 | 13 | 0.325 | 616 | 186 | 0.302 | 0.308 |
| 1977 | 42 | 9 | 0.214 | 595 | 183 | 0.308 | 0.273 |
| 1978 | 63 | 17 | 0.270 | 617 | 183 | 0.297 | 0.283 |

* AP statistics.

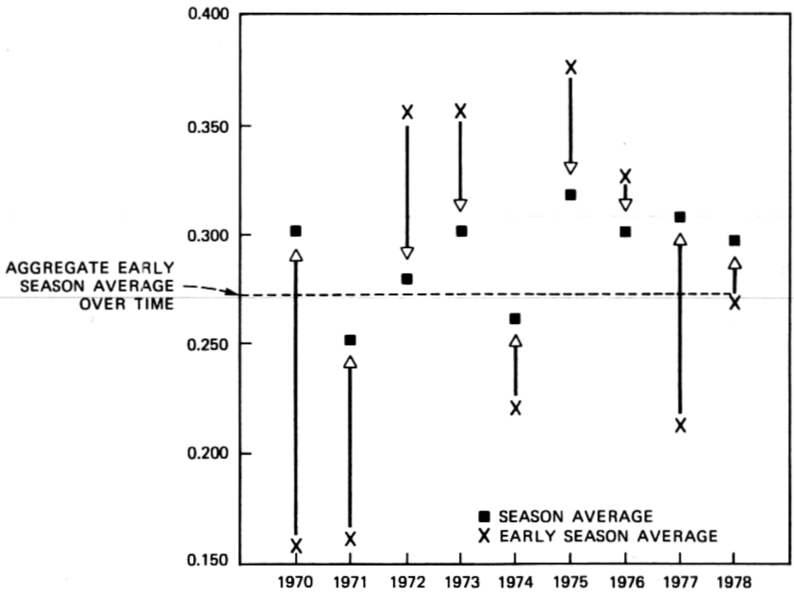


Fig. 3—Batting averages for Thurman Munson. For each year, the movement from the early season average (the sample) to the season average (the population) is always in the direction of the time average of the samples. This suggests strongly that by shrinking the samples towards their time average, one can obtain improved estimates of the populations.

the audit, we are interested in making a statistical inference each period about the current population. So our problem in Fig. 3 is to make a statistical inference each year about the season batting average using only the early season averages observed to date.

As an estimate, one would be tempted to use the maximum likelihood estimate, the early season average. But, in every year, the movement of the batting average from the early season to the season end is in the direction of the aggregate early season average over time. So, paradoxically, the early season averages from other years seem to be relevant to the current season average. It is clear from the data, that a better estimate of the season average is some kind of shrinkage of the early season average towards the aggregate early season average over time. And the amount of shrinkage can depend only on the available data—the early season averages.

What we really have here is a multivariate problem. We observe a nine-dimensional vector of observations whose mean is a vector of population characteristics, one of which we are particularly interested in. Stein[8] showed (for the normal distribution) that the maximum likelihood estimate of the vector is inadmissible. Why this is true manifests itself in baseball lore. A player that starts the season relatively hot, usually cools off; and a player that starts in a relative slump usually improves. This is due to the nature of sampling error. The hot player is usually partially lucky and the slumping player is usually partially unlucky.

QMP is based on the concepts illustrated by the Munson data. We saw in Fig. 1a that the Best Measure of the population index is between the current observed index and the estimated long-run process average.

The approach used for QMP is actually Bayesian empirical Bayes. The shrinkage factor used is a Bayes estimate of an optimal shrinkage factor. So the Best Measure has the form

$$W \begin{bmatrix} \text{estimated} \\ \text{process} \\ \text{average} \end{bmatrix} + (1 - W) \begin{bmatrix} \text{current} \\ \text{sample} \\ \text{index} \end{bmatrix},$$

where $W$ is a Bayes estimate of

$$\frac{[\text{sampling variance}]}{[\text{sampling variance}] + [\text{process variance}]}.$$

The bigger the sampling variance is, relative to the process variance, the more weight is put on the estimated process average.

There are two advantages to the Bayesian empirical Bayes approach over the approach in Ref. 10. One is that the weight, $W$, is always strictly between zero and one. This is because $W$ is a Bayes estimate of an unknown optimal weight, $\omega$, which has a nondegenerate posterior distribution on the interval [0, 1]. The approach taken in Ref. 10 is to use maximum likelihood estimates of $\omega$, which can be one; i.e., total shrinkage to the process average.

The second advantage is that an interval estimate of the current population index can be constructed from its posterior distribution. Most of the literature (e.g., Ref. 10) treats the estimation problem thoroughly, but it provides little guidance for the interval estimation problem.

The QMP algorithm is applied to the Munson data and the QMP estimates of the season averages are given in Table I. The sums of the absolute errors for the maximum likelihood estimates (April averages) and the QMP estimates are 0.603 and 0.331, respectively—a forty-five percent improvement. Notice that the QMP estimates for 1970 and 1971 are close to the April averages. This is because there was no history on Munson. The reduction in total absolute error for the years 1973 through 1978 was sixty-five percent, because of the benefit of history.

### *1.5 Objectives*

This paper is intended to document QMP. It contains the rationale for changing the rating system, a synopsis of QMP features, mathematical derivations of the rating formulas, the dynamics of QMP, the operating characteristics of QMP, many examples, and the QMP reporting format.

Readers who are interested only in the mathematics of QMP and how it relates to empirical Bayes, may skip Section II. Readers, who are not interested in the mathematical derivation of QMP, may skip Section IV.

## II. *T*-RATE SYSTEM

To understand the rationale behind QMP, one must first understand the $T$-rate system. From this we shall see where things have changed and where things have remained the same.

### *2.1 Finding defects*

The sampling methods along with the scope of inspection provide for a sample of units of count for each set of inspections. A unit of count is either a unit of product or a unit of a product's attribute such as solderless wrapped connections.

The result of conducting a set of inspections is a list of defects found and their descriptions. Frequently, underlying a defect is a variable measurement* that falls outside a range. QMP does not affect the process of finding defects.

### *2.2 Assessing defects*

The defects found sometimes occur in clusters for which the effect of the cluster is nonadditive; i.e., the effect is less than the sum of the effects of the individual defects occurring by themselves. In this case, the number of defects assessed for rating purposes is less than the number found. The defect assessment practices for the $T$-rate system evolved over a 50-year period, so these practices were based on a variety of criteria and engineering judgements. The defect assessment practices under QMP amount to a redesign of the practices using a single principle, which is described in Section 3.1.

### *2.3 Defect weighting and demerits*

The defects assessed are transformed into demerits or defectives or may remain as simple unweighted defects. In an audit based on demerits, each defect assessed is assigned a number of demerits: 100, 50, 10, or 1 for $A$, $B$, $C$, or $D$ weight defects, respectively. Guidelines for assigning demerit weights are contained in numerous general and

---

* For rating transmission characteristics of exchange area cable, some variables measurements are used directly without conversion to defects. We do not treat this case here.

special purpose demerit lists. The principles underlying these demerit lists are described by Dodge in Refs. 5 and 6. In an audit based on defectives, all defects found in a unit of product are analyzed to determine if the unit is considered defective. The assessment is either one or zero defectives. These transformations to demerits, defectives, or defects are not affected by QMP.

### 2.4 Quality standards

For any set of inspections, the quality engineers in the QAC have established quality standards. To do this, they considered audit scope, shop capability, field performance, economics, complexity, etc. The philosophy of standards is described in Ref. 3. For audits based on defects or defectives, the standards are expressed in defects or defectives per unit. For audits based on demerits, the standards are derived from fundamental defect per unit of count standards for $A$, $B$, $C$, $D$-type defects. In addition, we use Poisson as the standard distribution of the number of type $A$ defects (for example).

To make this clear, let's consider a simple example. Suppose in a sample of size $n$, there are $X_A, X_B, X_C, X_D$-type $A$, $B$, $C$, $D$ defects. The definition of standard quality is that $X_A, X_B, \cdots$ are independent and have Poisson distributions with means $n\lambda_A, n\lambda_B, \cdots$. The number of demerits in the sample is

$$D = 100X_A + 50X_B + 10X_C + X_D.$$

The mean and variance of $D$, given standard quality, are

$$\begin{aligned} E(D \mid S) &= 100(n\lambda_A) + 50(n\lambda_B) + \cdots \\ &= n[100\lambda_A + 50\lambda_B + 10\lambda_C + \lambda_D] \\ &= nU_s \\ V(D \mid S) &= (100)^2(n\lambda_A) + (50)^2(n\lambda_B) + \cdots \\ &= n[(100)^2\lambda_A + (50)^2\lambda_B + (10)^2\lambda_C + \lambda_D] \\ &= nC_s. \end{aligned}$$

The notation "$D \mid S$" reads "$D$ conditional on $S$."

Note that $U_s$ is the demerit per unit standard and $C_s$ is a variance per unit standard. These are the numbers that would be published in the official list of standards called the Master Reference list. The quality standards are not affected by QMP.

### 2.5 Rating classes and periods

For the purpose of reporting quality results to management, the products are grouped into rating classes. An example is: ESS No. 1

wired equipment, functional test, at Dallas.* The results of all the inspections associated with this rating class are aggregated over a time period called a *rating period.* A rating period is about six weeks long and there are eight rating periods per year. QMP does not affect the rating classes or periods.

### 2.6 The T-rate

The advantage of having quality standards is that observed quality results can be statistically compared to the standards. In the $T$-rate system, this is done with a statistic called the $T$-rate.

For a given rating class, let $Q$ denote the total number of defects, defectives, or demerits that are observed in all the inspections conducted on all the subproducts during a rating period. Because there are quality standards for each set of inspections on each product subclass, it is possible to compute the standard mean and variance of $Q$, denoted by $E(Q \mid S)$, $V(Q \mid S)$. The $T$-rate is

$$T\text{-rate} = \frac{E(Q \mid S) - Q}{\sqrt{V(Q \mid S)}}.$$

It measures the difference between the observed result and its standard in units of statistical standard deviation.

For each rating period, the $T$-rate is plotted in the control chart format shown in Fig. 1b. The control limits of $\pm 2$ are reasonable under the assumption that $Q$ has an approximate normal distribution. Then the standard distribution of $Q$ is the "standard normal," and excursions outside the control limits are rare under standard quality. For large audit sample sizes, this approximation follows from the central limit theorem. As we shall see, the approximation is poor for small sample sizes.

### 2.7 Reports, Below Normals and ALERTs

The fundamental reports to *WE* management are books of $T$-rate control charts for all rating classes. However, every rating period, a summary booklet is prepared. The summary consists of various aggregate quality performance indices and an exception report which lists rating classes that are having quality problems.

There are two kinds of exceptions: Below Normal (BN) and ALERT. These are based on statistical tests of the hypothesis that quality is at standard. The rules for BN and ALERT are based on six consecutive $T$-rates, $t_1, \cdots t_6$, where $t_6$ is the current $T$-rate. The rules use the

---

* Technically, this is called a scoring class in quality assurance documentation. Here, rating class means scoring class.

following runs criteria:

SCAN($S$): $t_1 < 0, \cdots, t_6 < 0$,

341 ($T$): $t_6 < -1$ and at least two of the set $\{t_3, t_4, t_5\}$ are less than $-1$.

Finally, the rules for BN and ALERT are:

*Below Normal* (BN): One of the following two conditions is satisfied:

(1) $t_6 < -3$

(2) $-3 \le t_6 < -2$ and at least one of the following three conditions hold:

(*i*) SCAN

(*ii*) 341

(*iii*) At least one of the set $\{t_2, t_3, t_4, t_5\}$ is less than $-2$.

ALERT: SCAN or 341 but not BN.

In Fig. 1b, examples of BN are 7806 and 7803. Examples of ALERT are 7808 and 7804.

Both the fundamental report formats and the rules for BN and ALERT are different under QMP

### *2.8 Pros and cons of the T-rate*

The advantage of the $T$-rate is its simplicity. It can be calculated manually. Exceptions can be identified by inspection. The fact that the $T$-rate has been used for so long is a testimonial to its advantages.

However, the $T$-rate does have problems.* The $T$-rate does not measure quality. A $T$-rate of $-6$ does not mean that quality is twice as bad as when the $t$-rate is $-3$. The $T$-rate is only a measure of statistical evidence with respect to the hypothesis of standard quality. This subtle statistical point is often misunderstood by report readers. Years of explanations have not cleared up the confusion.

Another problem is that the ALERT (SCAN and 341) rules are tests of hypothesis about quality trends, not current quality. Consider Fig. 2. You can assert that quality was probably substandard sometime between 7702 and 7707. You cannot, however, assert that quality is substandard in 7707. The QMP result for 7707 is normal.

In addition, the rules for ALERT and to some extent BN depend on attributes of past $T$-rates rather than their exact values. For example, five consecutive past $T$-rates at $-1.0$ are treated exactly like five consecutive $T$-rates at $-0.1$. This was done for statistical robustness. But statistical information is lost. There are no "outliers" in the audit data. Defects assessed were in the product. Many defects assessed

* Although the foundation of the $T$-rate system was laid by Shewhart[4] and Dodge,[5] the details are the results of contributions by many people over 50 years.

mean substandard quality at the time of assessment. It is possible that very unusual circumstances caused the defects. But it is intended that the audit flag such unusual circumstances.

The significance level of the $T$-rate hypothesis test depends on sample size and can be very large. Suppose that we have a simple test defect audit with a sample size of 32 units and a standard of 0.005 defects per unit. The expected number of defects is (32)(0.005) = 0.16. For one defect observed, the $T$-rate is

$$t = \frac{0.16 - 1}{\sqrt{0.16}} = -2.1.$$

So, every time there is a defect, the $T$-rate exceeds the control limit.

Now, assuming standard quality, the number of defects has a Poisson distribution with mean = 0.16. The Poisson probability of one or more defects is 0.15. So even when the standard is being met, there is a 15 percent chance of the $T$-rate dropping below −2.0. In statistical terms, we have a biased test (i.e., there is no reasonable upper bound on the significance level).

Clearly, it is not reasonable to take action every time the audit finds a defect. So special rules called modification treatments have evolved to handle cases like the one just described. Some of these modification treatments are statistically sound, others are not. This detracts from the desired objectivity of our quality rating.

In a sense, QMP is orthogonal to the $T$-rate. On the one hand, QMP cannot easily be computed manually. On the other hand, QMP does not have any of the disadvantages described above for the $T$-rate. The basic message of the QMP box chart (Fig. 1a) is unambiguous and exceptions can be identified by inspection.

## III. OVERVIEW OF QMP

As described in the introduction, QMP is the new way of conducting three of the audit ingredients: defect assessment, quality rating, and quality reporting. This section contains an overview of QMP. Mathematical derivations and detailed analyses of QMP are left for later sections.

### 3.1 Defect assessment practices

Defect assessment practices have two parts. Part one is a description of those situations where fewer defects are assessed than are found. Part two is a formula for the number of defects assessed.

In QMP, the principle for part one is: Normally all defects found in the quality assurance audits are assessed. Occasionally a cluster of two or more defects is found for which the seriousness of the cluster is less than the seriousness implied by individually assessing every defect in the cluster. Such a cluster shall be called *reducible*. Seriousness is measured from the customer's point of view. The audit attempts to measure seriousness as if the auditor is the customer. So if defects are found and corrected as a result of the audit, *no* adjustment in assessment is necessary. More specifically, a reducible cluster is a collection (on one audited unit) of

[1] dependent identical defects that the customer will
[2] almost surely discover in its entirety when a small part of the cluster is discovered and
[3] will correct or otherwise account for en masse, so that
[4] total seriousness is better represented by assessing $d_a$ defects (computed by the assessment formula), rather than the number found.

In [1], we use the word dependent in a statistical sense. Defects are dependent if they occur in a short interval of time and are systematically introduced by a common feature of the production process.

Ideally, the assessment associated with a reducible cluster of defects should depend on the situation. Over time, lists of reducible clusters and their assessments could be catalogued and added to the demerit lists. But, for now, there is no list of reducible clusters, so an assessment formula is needed.

For QMP, the assessment formula has the general form

$$d_a = \text{AN} + 1,$$

where AN stands for "Allowance Number." In turn, AN has the general form $\text{AN} = e + 3\sqrt{e}$, rounded down to an integer or to the closest integer, where $e$ can be interpreted as an expected number of defects. The computation of $e$ and the rounding depend on the audit. For some audits, tradition has prevailed, and for other audits, methods of computing $e$ were developed for QMP.

As an example, consider a single relay for which three contacts are defective (class B defect). The traditional method of computing $e$ for an apparatus audit is

$$e = (12)(0.005) = 0.06,$$

where 12 is the number of contacts in the relay and 0.005 is a traditional generic standard per unit of count for class-B defects. The traditional rounding is down, so $\text{AN} = (0.06) + 3\sqrt{0.06} = 0.79$, rounded down to zero. Hence, one class-B defect is assessed.

Another example is a reducible cluster of loose terminations found on a bay of equipment in a transmission installation performance audit. In this case, $e$ is just the quality standard for the bay in defects per unit, and the rounding is to the nearest integer.

As you have gathered by now, defect assessment is an art not a science. The principles and rules described here have empirical validity. In practice, they usually lead to reasonable assessments.

### 3.2 Equivalent defects and expectancy

A complicating factor in the analyses of audit results is that defects, defectives, and demerits are different. But, are they really? The answer is no; because, for statistical purposes, they can all be transformed into equivalent defects that have approximate Poisson distributions.

Suppose we have a quality measure $Q$ (total defects, defectives, or demerits). Let $E_s$ and $V_s$ denote the standard mean (called expectancy) and variance of $Q$. So the $T$-rate is $T = (E_s - Q)\sqrt{V_s}$.

Now define

$$X = \text{equivalent defects} = \frac{Q}{V_s/E_s}$$

and

$$\begin{aligned} e &= \text{equivalent expectancy} = \text{standard mean of } X \\ &= \frac{E_s}{V_s/E_s} = \frac{E_s^2}{V_s}. \end{aligned}$$

If all defects have Poisson distributions and are occurring at $\theta$ times the standard rate, then it can be shown that

$$E[X \mid \theta] = V(X \mid \theta) = e\theta;$$

hence, $X$ has an approximate Poisson distribution with mean $e\theta$.

As an example, consider the demerits case. The total number of demerits has the general form

$$D = \sum w_i X_i,$$

where the $w_i$'s are known weights and the $X_i$'s have Poisson distributions. Assume that the mean of $X_i$ is $e_i\theta$, where $e_i$ is the standard mean of $X_i$ and $\theta$ is the population quality expressed on an index scale. So $\theta = 2$ means that all types of defects are occurring at twice the rate expected.

The mean and variance of $D$ are

$$\begin{aligned} E(D) &= \sum w_i E(X_i) \\ &= \sum w_i(e_i\theta) \\ &= \theta E_s \end{aligned}$$

and

$$V(D) = \sum w_i^2 V(X_i)$$
$$= \sum w_i^2 (e_i \theta)$$
$$= \theta V_s,$$

where $E_s$ and $V_s$ are the standard mean and variance, respectively, of $D$.

The mean and variance of equivalent defects, $X$, are

$$E(X) = \frac{E(D)}{V_s/E_s}$$
$$= \frac{\theta E_s}{V_s/E_s}$$
$$= \theta e$$

and

$$V(X) = \frac{V(D)}{[V_s/E_s]^2}$$
$$= \frac{\theta V_s E_s^2}{V_s^2}$$
$$= \theta e.$$

The mean and variance of $X$ are equal; so, $X$ has an approximate Poisson distribution with mean $e\theta$. Of course, it is not exact; because, $X$ is not always integer valued. But, this Poisson approximation for equivalent defects is better than the normal approximation implied by the $T$-rate system. It is the Poisson approximation in QMP that obviates the need for the modification treatments discussed in Section 2.8.

A similar analysis works approximately for the defectives case. So, any aggregate of demerits, defectives, or defects can be transformed into equivalent defects. Just use the standard expectancy and variance as illustrated above for demerits.

### 3.3 Statistical foundations of QMP

The algorithm used for computing the QMP box chart shown in Fig. 1a was derived from a Bayesian analysis of a particular statistical model. In this Section we describe the model and put it in perspective. This will provide an appreciation for how the box charts can be interpreted and why they are a useful management tool.

#### 3.3.1 QMP model

For rating period $t$, let $x_t$ = equivalent defects in the audit sample, $e_t$ = equivalent expectancy of the audit sample, $\theta_t$ = population index, as defined in Section 3.2. Based on the discussion in Section 3.2, we assume that the conditional disribution of $x_t$ given $\theta_t$ is Poisson with mean $e_t\theta_t$; i.e.,

$$x_t \mid \theta_t \sim \text{Poisson}(e_t\theta_t).$$

In Fig. 3 we see that the season average varies from year to year. Some of that variation is due to the fact that the season is itself a sample from a conceptual infinite population of at bats. The rest of the variation is due to changes in ability, competition, etc., that are caused by numerous factors that may or may not be identifiable. The important concept is that the time series of season averages is a stochastic process. For QMP we assume that the time series of $\theta_t$'s is an unknown stochastic process.

For reasons that are partly statistical and partly administrative, we have decided to restrict our use of past data to five periods. The main administrative reason is that the $T$-rate system used the past five periods. So all of the $T$-rate administrative rules that dealt with missing data and reinitialization of rating classes can be used in QMP. Statistically, QMP works as well for six periods as it does for eight periods (one year).

A consequence of using only six periods of data is that no useful inference can be made about possible complex structure in the stochastic process of $\theta_t$'s. So we assume simply that the $\theta_t$'s are a random sample from an unknown distribution called the *process distribution*. Furthermore, six observations are not enough to make fine inferences about the family of this unknown distribution. So for mathematical simplicity we assume it to be a gamma distribution with unknown mean $= \theta$ and variance $= \gamma^2$ (Appendix A); i.e.,

$$\theta_t \overset{\text{iid}}{\sim} \text{Gamma}\left(\frac{\theta^2}{\gamma^2}, \frac{\gamma^2}{\theta}\right),$$

$$t = 1, \cdots, T(\text{current period}).$$

The parameters $\theta^2/\gamma^2$ and $\gamma^2/\theta$ are the shape and scale parameters of the gamma distribution. We use the names

$$\theta = \text{process average},$$

$$\gamma^2 = \text{process variance}.$$

This choice of a unimodal distribution reflects our experience that usually many independent factors affect quality; so there is a central limit theorem effect.

We are assuming that the process average is unknown but fixed. In reality, it may be changing. We handle this by using a moving window of six periods of data. But this treats the past data symmetrically. An alternative would be some kind of exponential smoothing or Kalman filtering. My colleague M. S. Phadke is developing a generalization to QMP based on a random walk model for the process average.

The model so far is an empirical Bayes model.[10] The parameter of interest is the current population index, $\theta_T$, which has a distribution called the process distribution. Bayesians would call it the prior distribution if it were known. But we must use all the data to make an inference about the unknown process distribution. So, the model is called empirical Bayes.

Efron and Morris[10] take a classical approach to the empirical Bayes model. They use classical methods of inference for the unknown process distribution. QMP is based on a Bayesian approach to the empirical Bayes model. Each product has its own process mean and variance. These vary from product to product. By analyzing many products, we can model this variation by a prior distribution for $(\theta, \gamma^2)$.

Summarizing, our model is

$$x_t \mid \theta_t \sim \text{Poisson}(e_t\theta_t), \qquad t = 1, \cdots, T,$$

$$\theta_t \overset{\text{iid}}{\sim} \text{Gamma}\left(\frac{\theta^2}{\gamma^2}, \frac{\gamma^2}{\theta}\right), \ (\theta, \gamma^2) \sim \text{prior distribution } \rho(\theta, \gamma^2).$$

For now, $\rho(\theta, \gamma^2)$ remains general.

This is a full Bayesian model. It specifies the joint distribution of all variables. The quality rating in QMP is based on the posterior distribution of $\theta_T$ given $\mathbf{x} = (x_1, \cdots, x_T)$.

### 3.3.2 The model in perspective

Quality rating in QMP is based on posterior probabilities given the audit data. Of course these probabilities depend on the model. But how do we know the model is right?

It is important to understand that we are not doing data analysis with QMP. In data analysis, each set of data is treated uniquely. Probabilities cannot be computed. Objective decisions cannot be made.

A requirement of quality rating is a specific rule that defines quality exceptions and a figure of merit (e.g., a probability) associated with an exception. A statistical model provides both. QMP could have been based on a more elaborate model. Our model represents a compromise between simplicity and believability.

So our exception decisions are at least consistent with one simple model of reality. The probabilities are conditional on that model. Otherwise, they can only be interpreted as figures of merit.

We have imbedded the simple hypothesis of a Poisson distribution with a standard mean into a class of alternatives. The alternatives are

Poisson distributions with nonstandard means. Much more complicated alternatives can be included: e.g., the class of negative binomial distributions, and our probabilities would change a little. But QMP has achieved a kind of empirical validity. The exceptions being identified are accepted by the managers being rated. And for the products declared normal, there is a model (i.e., our model) that affords the standard hypothesis some credence.

### 3.3.3 Posterior distribution of current quality

We show in Section IV that it is computationally impractical to derive the exact posterior distribution of $\theta_T$. The best we can do is approximate the posterior mean and variance of $\theta_T$.

The posterior mean and variance of $\theta_T$ are derived in Section IV. The posterior mean is

$$\begin{aligned}\hat{\theta}_T &= E(\theta_T \mid \mathbf{x}) \\ &\doteq \hat{\omega}_T \hat{\theta} + (1 - \hat{\omega}_T) I_T,\end{aligned}$$

where

$$\begin{aligned}\hat{\theta} &\doteq E(\theta \mid \mathbf{x}), \\ \hat{\omega}_T &\doteq E(\omega_T \mid \mathbf{x}), \\ \omega_T &= \frac{\theta/e_T}{\theta/e_T + \gamma^2}.\end{aligned}$$

The posterior mean, $\hat{\theta}_T$, is a weighted average between the estimated process average, $\hat{\theta}$, and the defect index, $I_T$, of the current sample. It is the dynamics of the weight, $\omega_T$, that makes the Bayes estimate work so well. For any $t$, the sampling variance of $I_t$ is

$$\begin{aligned}V(I_t \mid \theta_t) &= V\left(\frac{x_t}{e_t} \,\middle|\, \theta_t\right) \\ &= \frac{1}{e_t^2} V(x_t \mid \theta_t) \\ &= \frac{1}{e_t^2} (e_t \theta_t) \\ &= \theta_t / e_t.\end{aligned}$$

The expected value of this is

$$E[\theta_t / e_t] = \theta / e_t.$$

So the weight, $\omega_T$, is

$$\frac{[\text{expected sampling variance}]}{[\text{expected sampling variance}] + [\text{process variance}]}.$$

If the process is relatively stable, then the process variance is relatively small and the weight is mostly on the process average; but if the process is relatively unstable, then the process variance is relatively large and the weight is mostly on the current sample index. The reverse is true of the sampling variance. If it is relatively large (e.g., small expectancy), then the current data is weak and the weight is mostly on the process average; but if the sampling variance is relatively small (e.g., large expectancy), then the weight is mostly on the current sample index. In other words, $\omega_T$, is a monotonic function of the ratio of expected sampling variance to process variance.

The posterior variance of $\theta_T$ is

$$V_T \doteq (1 - \hat{\omega}_T)\hat{\theta}_T/e_T + \hat{\omega}_T^2 V(\theta \mid \mathbf{x}) + (\hat{\theta} - I_T)^2 V(\omega_T \mid \mathbf{x}).$$

If the process average and variance were known, then the posterior variance of $\theta_T$ would be $(1 - \omega_T)\hat{\theta}_T/e_T$ (Appendix B). So the first term is just an estimate of this. But since the process average and variance are not known, the posterior variance has two additional terms. One contains the posterior variance of the process average and the other contains the posterior variance of the weight.

The first term dominates. A large $\hat{\omega}_T$ (relatively stable process), a small $\hat{\theta}_T$ (good current quality), and a large $e_T$ (large audit) all tend to make the posterior variance of $\theta_T$ small (the box chart short).

If $\hat{\omega}_T$ is small, then the second term is negligible. This is because the past data is not used much, so the uncertainty about the process average is irrelevant.

If the current sample index is far from the process average, then the third term can be important. This is because outlying observations add to our uncertainty as to what is happening.

If the process average and variance were known, then the posterior distribution would be gamma (Appendix B). So we approximate the posterior distribution with a gamma fitted by the method of moments. The parameters of the fitted gamma are

$$\begin{aligned} \alpha &= \text{shape parameter} \\ &= \hat{\theta}_T^2/V_T, \\ \tau &= \text{scale parameter} \\ &= V_T/\hat{\theta}_T, \end{aligned}$$

and the posterior cumulative distribution function is

$$\Pr\{\theta_T \le y \mid \mathbf{x}\} = G_\alpha(y/\tau)$$

(Appendix A).

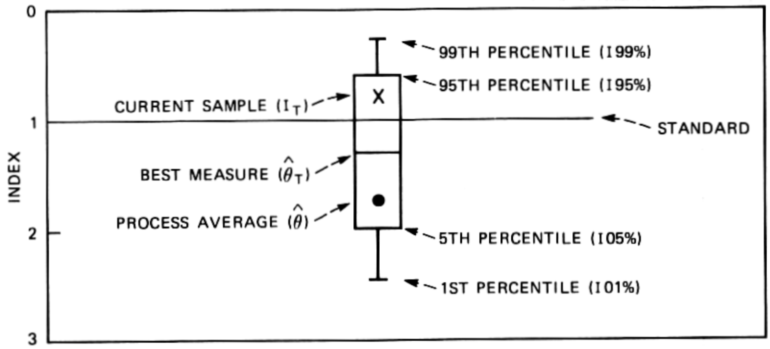


Fig. 4—QMP box and whisker chart. This is a graphical representation of the posterior distribution for current production given the six most recent periods of audit data. The whiskers display the 99th and 1st percentiles and the box displays the 95th and 5th percentiles. The Best Measure is the posterior mean or Bayes estimate. It is a weighted average of the process average ("dot") and the current sample ("X"). The weight is the ratio of sampling variance to total variance. If all the variance is due to sampling, then the production is stable and the process average is the Best Measure of current quality. If the sampling variance is zero, then the current sample is the Best Measure.

### 3.4 QMP reports

#### 3.4.1 QMP box chart

The QMP box and whisker chart is shown in Fig. 4. $I99\%, \cdots, I01\%$ are defined by

$$1 - G_\alpha(I99\%/\tau) = 0.99,$$
$$\vdots$$
$$1 - G_\alpha(I01\%/\tau) = 0.01.$$

So, e.g., *a posteriori*, there is a 99 percent chance that $\theta_T$ is larger than $I99\%$.

#### 3.4.2 QMP Below Normal and ALERT definitions

In QMP, a rating class is Below Normal (BN) if

$$I99\% > 1;$$

i.e., the posterior probability that the product is substandard exceeds 0.99. Substandard means $\theta_T > 1$. A rating class is on ALERT if

$$I99\% \leq 1 < I95\%;$$

i.e., the posterior probability that the product is substandard exceeds 0.95 but not 0.99.

These definitions are illustrated graphically in Fig. 5, which is oriented like the location summary in Fig. 6.

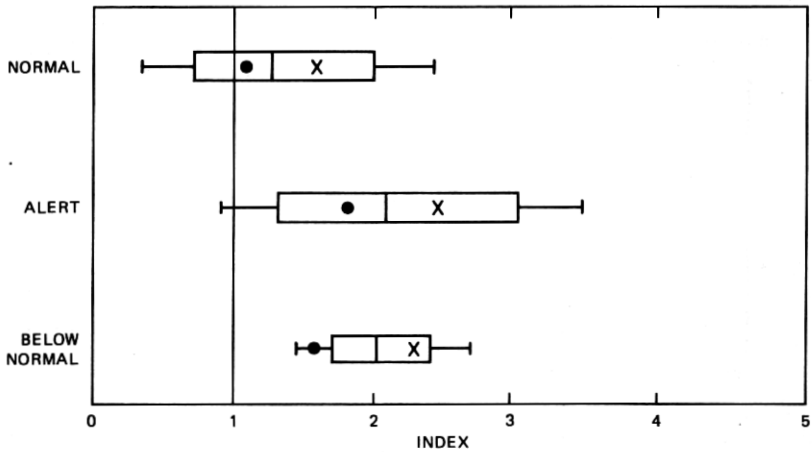


Fig. 5—QMP exceptions. Below Normal means that the probability of substandard quality exceeds 0.99. For ALERT, the probability exceeds 0.95 but not 0.99. Normal is not a quality exception.

### 3.4.3 QMP report formats

There are two report formats for QMP results. One is a time series of box charts illustrated in Fig. 1a. The estimated process averages are joined. The other is a location summary for the current rating period. This is illustrated in Fig. 6. It orders the rating classes by Best Measure for the current period. Another ordering that will be used is by rating class name.

Western Electric, Bell Laboratories, and American Telephone and Telegraph management will receive all QMP results. Operating company management will receive QMP results on those rating classes that are of direct interest to them. Examples of results provided to the operating companies are the quality of repaired telephone sets and installed switching systems.

### 3.5 Advantages of QMP

Many of the advantages of QMP relate to the disadvantages of the *T*-rate system (Section 2.8). QMP provides a direct measure of quality. If a rating class is Below Normal, one can tell how bad the quality is. QMP uses past and current data to make an inference about current quality not past quality. If a rating class is on ALERT, then it is over 95 percent probable that there is a quality problem *now.* QMP does not use runs criteria, but uses the actual equivalent defects observed. This provides more statisxtical efficiency and therefore shorter interval estimates. QMP is robust against statistical "jitter." It does not overreact to a few defects. Consequently, there is no need for special

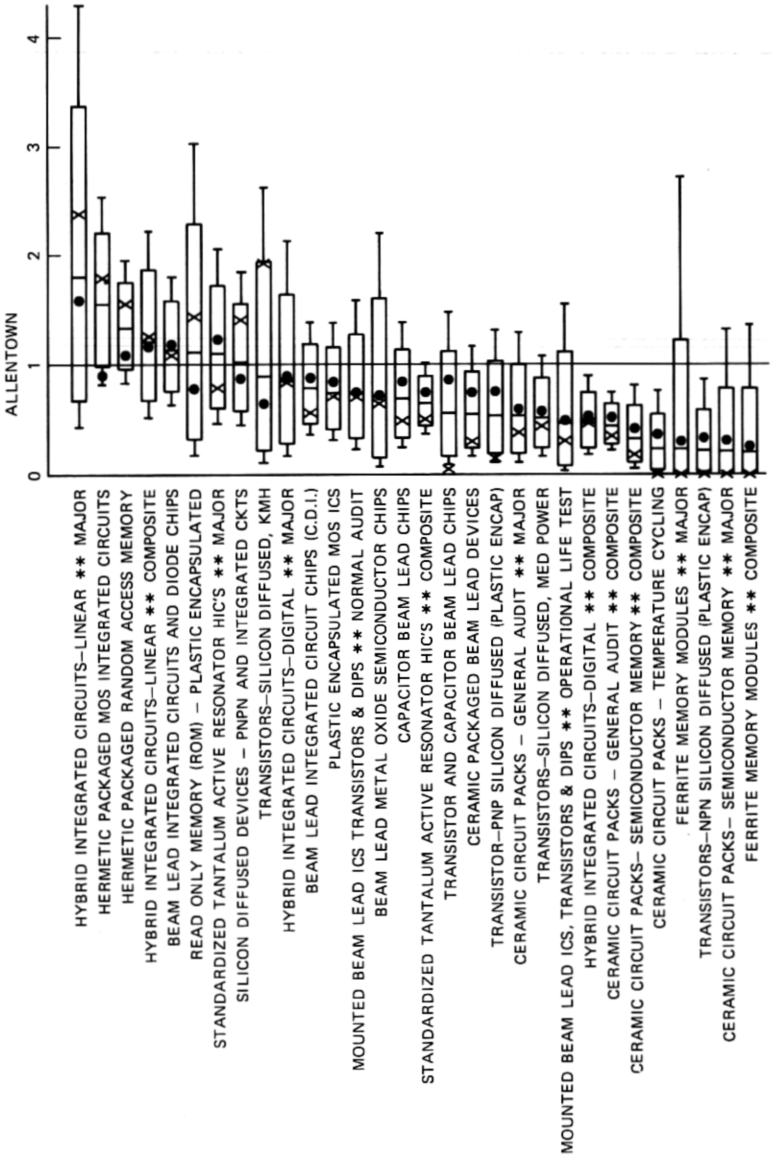


Fig. 6—QMP location summary.

modification treatments. This way we retain statistical objectivity conditional on our model.

Another advantage compared to the $T$-rate is that QMP provides a lower producer's risk and consequently a more accurate list of exceptions. This is supported by data presented in Section VI.

Finally, QMP will allow us to unify our reporting to Bell System management. In the past, the $T$-rate statistic did not meet the needs of the operating companies; so, we developed a collection of special reports for the operating companies. Since the QMP report format does meet operating company needs, the relevant subset of all the results is a useful report.

## IV. MATHEMATICAL DERIVATION OF QMP

### 4.1 Exact solution

We are interested in the posterior distribution of $\theta_T$ given $\mathbf{x}$, for the model described in Section 3.3.1. Now $\Pr\{\theta_T \le y \mid \mathbf{x}\} = \int_0^\infty \int_0^\infty \Pr\{\theta_T \le y \mid \theta, \gamma^2, x_T\}\rho(\theta, \gamma^2 \mid \mathbf{x})\, d\theta\, d\gamma^2$, where $\rho(\theta, \gamma^2 \mid \mathbf{x})$ is the posterior distribution of $\theta$, $\gamma^2$ given $\mathbf{x}$.

From Appendix B, we know that the distribution of $\theta_T$ given $\theta$, $\gamma^2$, and $x_T$ is gamma; so, $\Pr\{\theta_T \le y \mid \theta, \gamma^2, x_T\}$ can be expressed in terms of an incomplete gamma function.

By Bayes theorem,

$$\rho(\theta, \gamma^2 \mid \mathbf{x}) = \frac{\rho(\theta, \gamma^2)L(\theta, \gamma^2)}{\displaystyle\int_0^\infty \int_0^\infty \rho(\theta, \gamma^2)L(\theta, \gamma^2)\, d\theta\, d\gamma^2},$$

where $\rho(\theta, \gamma^2)$ is the prior density of $\theta$, $\gamma^2$ and $L(\theta, \gamma^2)$ is the likelihood function. Since $x_t$ given $\theta_t$ is Poisson and $\theta_t$ given $\theta$, $\gamma^2$ is gamma, it follows that $x_t$ given $\theta$, $\gamma^2$ is negative binomial; hence,

$$L(\theta, \gamma^2) = \prod_{t=1}^{T} L_t(\theta, \gamma^2),$$

where

$$L_t(\theta, \gamma^2) = \frac{\Gamma(x_t + \theta^2/\gamma^2)}{x_t!\,\Gamma(\theta^2/\gamma^2)}[1 - \omega_t(\theta, \gamma^2)]^{x_t}[\omega_t(\theta, \gamma^2)]^{\theta^2/\gamma^2},$$

$$\omega_t(\theta, \gamma^2) = \frac{\theta/e_t}{\theta/e_t + \gamma^2}.$$

So the posterior distribution of $\theta_T$ is a complex triple integral that has to be inverted to compute the QMP box chart. The posterior mean and variance of $\theta_T$ can be expressed in terms of several double integrals. There are more than 1,000 rating classes that have to be analyzed each period, so computational efficiency is important. This is why we have developed an efficient heuristic solution to the problem.

### 4.2 Empirical priors for process parameters

It is clear from Section 4.1 that prior distributions for $\theta$ and $\gamma^2$ are needed. In the fourth rating period of 1979, we applied an earlier version of the QMP algorithm to over 1,000 rating classes. This provided over 1,000 estimates of $\theta$ and $\gamma^2$ and empirical distributions of these estimates. The empirical mean and variance of the $\theta$ estimates were 0.75 and 0.17, respectively. The empirical mean, variance, and mode of the $\gamma^2$ estimates were 0.28, 0.19, and 0.05, respectively.

In the remainder of Section IV, we use 1 as a mean value of $\theta$ instead of 0.75. This is because 1 is the desired standard value that minimizes first cost plus maintenance costs. Under QMP, the shops will be able to operate on the average closer to 1, because the producer's risk (see Section 6.2) is smaller than for the $T$-rate. Also, more defects are assessed under QMP than for the $T$-rate (see Sections 2.2 and 3.1).

### 4.3 Posterior mean of current quality

For the model described in Section 3.3.1,

$$\begin{aligned}\hat{\theta}_T &= E(\theta_T \mid \mathbf{x}) \\ &= E[E(\theta_T \mid \theta, \gamma^2, \mathbf{x}) \mid \mathbf{x}].\end{aligned}$$

Conditioning on $\theta$ and $\gamma^2$ means that the process distribution is known. So by Theorem B.1 in Appendix B,

$$\hat{\theta}_T = E[\omega_T \theta + (1 - \omega_T) I_T \mid \mathbf{x}]. \tag{1}$$

To calculate this posterior expectation exactly requires a double integral. But a posterior expectation, $E[\cdot \mid \mathbf{x}]$, can be viewed as an estimate of the operand "$\cdot$", because it is the Bayes estimate. So all we need are estimates of $\omega_T$ and $\theta$.

#### 4.3.1 Moment estimates of process parameters

As argued in Section 4.1, given $\theta$ and $\gamma^2$, $x_t$ has a negative binomial distribution. We show in Appendix D, eqs. (56) and (58), that

$$\begin{aligned}E(I_t) &= \theta, \\ E(Y_t) &= \gamma^2,\end{aligned}$$

where

$$Y_t = (I_t - \theta)^2 - I_t / e_t. \tag{2}$$

So we have many independent estimates of $\theta$ and $\gamma^2$. A general method of combining independent estimates of parameters is a weighted average, where the weights are proportional to the reciprocal of the variances of the individual independent estimates. Such estimates of $\theta$ and $\gamma^2$ are

$$\bar{\theta} = \sum_t p_t I_t, \tag{3}$$

$$\overline{\gamma^2} = \sum_t q_t Y_t, \tag{4}$$

where

$$\sum_t p_t = \sum_t q_t = 1$$

and

$$p_t \propto 1/V(I_t),$$

$$q_t \propto 1/V(Y_t).$$

Notice that $Y_t$ depends on $\theta$. So in the application, we replace $\theta$ by an estimate.

Now $V(I_t)$ and $V(Y_t)$ depend on the unknown parameters $\theta$ and $\gamma^2$. The important consideration in setting the weights $p_t$ and $q_t$ is their general behavior as $e_t$ varies. So for simplicity (to avoid iteration), we choose $\theta = 1$ and $\gamma^2 = 1/4$, which are empirically-determined mean values of these process parameters (see Section 4.2).

In Appendix D, we derive formulas for $V(I_t)$ and $V(Y_t)$. Plugging $\theta = 1$ and $\gamma^2 = 1/4$ into eqs. (56) and (59) yields

$$p_t \propto f_t = \frac{1}{V(I_t)} = 1\Big/\left(\frac{1}{e_t} + \frac{1}{4}\right) = \frac{e_t}{1 + e_t/4}, \tag{5}$$

$$q_t \propto g_t = \frac{1}{V(Y_t)} = 1\Big/\left[\frac{2.5}{e_t^2} + \frac{1.5}{e_t} + 0.22\right]$$

$$= \frac{e_t^2}{2.5 + 1.5e_t + (0.22)e_t^2}. \tag{6}$$

Note that for small $e_t$, $f_t \propto e_t$; but for large $e_t$, the $f_t$'s and therefore the weights, $p_t$'s, are all about equal. This is because for any large $e_t$, $I_t \doteq \theta_t$ and we are trying to estimate the average of the $\theta_t$'s.

#### 4.3.2 Bayes estimate of the process average

In the case $I_t \equiv 0$ for all $t$, there is a problem with the estimate $\bar{\theta}$. If we plug $\bar{\theta} = 0$ into (1), then $\hat{\theta}_T = 0$. But $\hat{\theta}_T$ is a posterior mean of a positive parameter, so it cannot be zero. The correct method of handling this problem is to start with a proper prior distribution on the process average, $\theta$. But then the mathematics and the computations become complicated.

So we assert that we have prior information that is equivalent to observing some "prior data," $x_0$ and $e_0$. Then a Bayes type estimate has the form

$$\hat{\theta} = \sum_{t=0}^{T} p_t I_t, \tag{7}$$

which has the same form as the moment estimate, $\bar{\theta}$, but uses all the data including the "prior data."

To choose values for $x_0$ and $e_0$, consider $T = 1$. A generic form of a Bayes type estimate of $\theta$ is

$$wE(\theta) + (1 - w)I_1,$$

where

$$w = \frac{V(I_1)}{V(I_1) + V(\theta)}.$$

Setting this generic form equal to (7) yields

$$E(\theta) = x_0/e_0,$$

$$V(\theta) = 1/e_0 + 1/4.$$

From Section 4.2, $E(\theta) = 1$; and we conservatively choose $V(\theta)$ to be 1.25 (we do not want our prior observations of $\theta$ estimates to preclude large future values of $\theta$). This implies $x_0 = e_0 = 1$.

#### *4.3.3 Bayes estimate of weight*

Now define an estimate of $\gamma^2$ analogous to (7),

$$\hat{\gamma}_1^2 = \sum_{t=0}^{T} q_t(I_t - \hat{\theta})^2 - \sum_{t=0}^{T} q_t(I_t/e_t),$$

as suggested by eqs. (2) and (4).* The first term is the total variance about the process average and the second term is a weighted average of estimated sampling variances. [Recall from Section 3.3.3 that $V(I_t \mid \theta_t) = \theta_t/e_t$, which can be estimated by $I_t/e_t$.] We denote the average sampling variance by

$$\sigma^2 = \sum_{t=0}^{T} q_t(I_t/e_t). \tag{8}$$

The problem with $\hat{\gamma}_1^2$ as an estimate of process variance is that it can be negative. To solve this problem, we use the results in Appendix C. Assume $\sigma^2$ is a known constant, and define the unknown weight as

$$\omega = \frac{\sigma^2}{\sigma^2 + \gamma^2}.$$

* Note that we treat the "prior data" as real data.

To apply Appendix C, we must find a statistic, SS, and a degree of freedom, df, so that, approximately,

$$\frac{\omega}{\sigma^2}\,(\mathrm{SS}) \sim \chi^2_{\mathrm{df}}.$$

Originally, we just assumed approximate normality of $I_t$ and took SS $= (\mathrm{df}+1)\sum_{t=0}^{T} q_t(I_t - \hat{\theta})^2$ and df $= T$. But we found unusual sets of data for which the number of defects allowed (before declaration of Below Normal) was a decreasing function of expectancy in short ranges of small expectancy. We dubbed this the "QMP wiggle."

To solve this problem, we approximate the sampling distribution of

$$Z = \sum_{t=0}^{T} q_t(I_t - \hat{\theta})^2$$

by a scaled chi-square with degrees of freedom deduced by the method of moments.

Let

$$Z_1 = \sum_{t=0}^{T} q_t(I_t - \theta)^2,$$

and try an SS of the form

$$uZ,$$

where $u$ is an unknown constant. The two moment equations that have to be satisfied are

$$E\left[\frac{\omega}{\sigma^2}\,(uZ)\right] = \mathrm{df}, \tag{9}$$

$$\frac{E^2[(\omega/\sigma^2)(uZ)]}{V[(\omega/\sigma^2)(uZ)]} = \frac{E^2[\chi^2_{\mathrm{df}}]}{V[\chi^2_{\mathrm{df}}]}.$$

And the second equation is

$$\frac{2E^2[Z]}{V[Z]} = \mathrm{df}. \tag{10}$$

Inspired by well-known normal theory, we use the approximations

$$E(Z) = \left(\frac{\mathrm{df}}{\mathrm{df}+1}\right) E(Z_1), \tag{11}$$

$$\frac{2E^2(Z)}{V(Z)} = -\frac{2E^2(Z_1)}{V(Z_1)} - 1. \tag{12}$$

Now using eqs. (11) and (56), eq. (9) becomes

$$E\left[\frac{\omega}{\sigma^2}\,(uZ)\right] = \frac{\omega}{\sigma^2}\,u\left(\frac{\mathrm{df}}{\mathrm{df}+1}\right)\sum q_t(\gamma^2 + \theta/e_t)$$

$$\doteq \frac{\omega}{\sigma^2} u\left(\frac{\text{df}}{\text{df} + 1}\right) (\gamma^2 + \sigma^2)$$

$$= u\left(\frac{\text{df}}{\text{df} + 1}\right)$$

$$= \text{df};$$

hence,

$$u = \text{df} + 1.$$

As for eq. (12), the mean and variance of $Z_1$ depend on $\theta$ and $\gamma^2$. So to avoid iteration, we now select $\theta = 1$ and $\gamma^2 = 0$,* which were empirically determined in Section 4.2. Then by eqs. (12), (56), and (57),

$$\frac{2E^2[Z]}{V[Z]} = \frac{2[\sum q_t(1/e_t)]^2}{\sum q_t^2(1/e_t^3 + 2/e_t^2)} - 1$$
$$= \text{df}. \tag{13}$$

So our statistic ss is

$$\text{ss} = (\text{df} + 1) \sum q_t(I_t - \hat{\theta})^2,$$

where df is given by eq. (13).

Now apply the Corollary to Theorem C.1 in Appendix C to get

$$\omega \mid \text{ss} \sim C - \text{Gamma}\left(a, \frac{\sigma^2}{b}, 1\right),$$

where

$$a = a_0 + \frac{\text{df}}{2}, \qquad b = b_0 + \frac{\text{ss}}{2}. \tag{14}$$

Define

$$S^2 = b/a$$
$$= \frac{2b_0 + (\text{df} + 1) \sum_{t=0}^{T} q_t(I_t - \hat{\theta})^2}{2a_0 + \text{df}}, \tag{15}$$

$$R = S^2/\sigma^2. \tag{16}$$

Now apply Theorem C.2 in Appendix C to get

$$E(\omega \mid \mathbf{x}) = \frac{1}{\text{RF}} = \frac{\sigma^2}{\text{FS}^2}, \tag{17}$$

* The choice of $\gamma^2 = 0$ here may seem inconsistent with choice of $\gamma^2 = 1/4$ in the definition of $f_t$ and $g_t$. There it was necessary to take a positive value (the empirical mean across products) of $\gamma^2$ to get the correct behavior for large $e_t$. Here it was not necessary, so for simplicity, we took the approximate empirical mode across products, $\gamma^2 = 0$.

$$V(\omega \mid \mathbf{x}) = G. \tag{18}$$

To determine the parameters $(a_0, b_0)$ of the prior distribution of $\omega$, we first develop an empirical distribution of estimated $\omega$'s across many rating classes, which have a mean and variance of 0.6 and 0.03, respectively. To be conservative, we inflate the variance, shrink the mean, and select the prior mean and variance of $\omega$ to be 0.55 and 0.045, respectively. The parameters $a_0$ and $b_0$ are then solutions to (see Appendix C)

$$\frac{1}{R_0 F} = 0.55,$$

$$G = 0.045,$$

where $F$ and $G$ are defined in Theorem C.2 in terms of $a_0$ and $R_0 = b_0/a_0\sigma^2$. A numerical analysis yields

$$a_0 = 4.5,$$

$$R_0 = b_0/a_0\sigma^2 = 1.6,$$

or

$$a_0 = 4.5, \tag{19}$$

$$b_0 = (7.2)\sigma^2. \tag{20}$$

Now we define an estimate, $\hat{\gamma}^2$, of the process variance by

$$E(\omega \mid \mathbf{x}) = \frac{\sigma^2}{\sigma^2 + \hat{\gamma}^2}.$$

So by eq. (14)

$$\hat{\gamma}^2 = FS^2 - \sigma^2$$

$$= (FR - 1)\sigma^2. \tag{21}$$

This is our improvement of the moments estimate. The inflation factor $F$ prevents $\hat{\gamma}^2$ from being negative or zero. It can be shown that if $R$ is large, $F$ is approximately one; but if $R$ is small, $FR - 1$ is positive and $F$ is large.

We are now in a position to estimate $\omega_T$ by

$$\hat{\omega}_T = \frac{\sigma_T^2}{\sigma_T^2 + \hat{\gamma}^2}, \tag{22}$$

where

$$\sigma_T^2 = \hat{\theta}/e_T, \tag{23}$$

and our approximation to (1) is

$$\hat{\theta}_T = \hat{\omega}_T\hat{\theta} + (1 - \hat{\omega}_T)I_T. \tag{24}$$

### 4.4 Posterior variance of current quality

For the model described in Section 3.3.1,

$$V_T = V(\theta_T | \mathbf{x}) = E[V(\theta_T | \theta, \gamma^2, \mathbf{x}) | \mathbf{x}] + V[E(\theta_T | \theta, \gamma^2, \mathbf{x}) | \mathbf{x}].$$

Conditioning on $\theta$ and $\gamma^2$ amounts to the process distribution being known. So by Theorem B.1, in Appendix B,

$$V_T = E[(1 - \omega_T)E(\theta_T | \theta, \gamma^2, \mathbf{x})/e_T | \mathbf{x}] + V[\omega_T(\theta - I_T) | \mathbf{x}].$$

Conditioning on $\gamma^2$ in the second term yields

$$\begin{aligned} V_T = {} & E[(1 - \omega_T)E(\theta_T | \theta, \gamma^2, \mathbf{x})/e_T | \mathbf{x}] \\ & + E[V[\omega_T(\theta - I_T) | \gamma^2, \mathbf{x}] | \mathbf{x}] \\ & + V[E[\omega_T(\theta - I_T) | \gamma^2, \mathbf{x}]\mathbf{x}], \end{aligned} \tag{25}$$

so the posterior variance has three components.

#### 4.4.1 First component

The first component is approximated by regarding the posterior expectation operator as an estimation operator, and it is

$$(1 - \hat{\omega}_T)\hat{\theta}_T/e_T. \tag{26}$$

#### 4.4.2 Second component

To approximate the second component, we first approximate $V[\omega_T(\theta - I_T) | \gamma^2, \mathbf{x}]$. Since $\omega_T$ depends primarily on $\gamma^2$ and $e_T$, we shall consider $\omega_T$ a constant. So

$$V[\omega_T(\theta - I_T) | \gamma^2, \mathbf{x}] \doteq \omega_T^2 V(\theta | \gamma^2, \mathbf{x}). \tag{27}$$

We use the approximation

$$V(\theta | \gamma^2, \mathbf{x}) \doteq V(\hat{\theta} | \gamma^2, \theta = \hat{\theta}). \tag{28}$$

Now by eq. (57),

$$\begin{aligned} V(\hat{\theta} | \gamma^2, \theta) &= V(\textstyle\sum p_t I_t | \gamma^2, \theta) \\ &= \textstyle\sum p_t^2 V(I_t | \gamma^2, \theta) \\ &= \textstyle\sum p_t^2[\gamma^2 + \theta/e_t]. \end{aligned} \tag{29}$$

Plugging eqs. (28) and (29) into eq. (27) yields

$$V[\omega_T(\theta - I_T) | \gamma^2, \mathbf{x}] \doteq \omega_T^2 \sum_{t=0}^{T} p_t^2[\gamma^2 + \hat{\theta}/e_t].$$

Again, treating the posterior expectation operator as an estimation operator, we get for the second component of eq. (25)

$$\hat{\omega}_T^2 \sum_{t=0}^{T} p_t^2[\hat{\gamma}^2 + \hat{\theta}/e_t]. \tag{30}$$

#### 4.4.3 Third component

For the third component in eq. (25), we first approximate $E[\omega_T(\theta - I_T) \mid \gamma^2, \mathbf{x}]$ by $\bar{\omega}_T(\hat{\theta} - I_T)$, where

$$\bar{\omega}_T = \frac{\hat{\theta}/e_T}{\hat{\theta}/e_T + \gamma^2}.$$

So the third component in eq. (25) is

$$(\hat{\theta} - I_T)^2 V(\bar{\omega}_T \mid \mathbf{x}). \tag{31}$$

If we define

$$r_T = \frac{\hat{\theta}/e_T}{\sigma^2},$$

$$\omega = \frac{\sigma^2}{\sigma^2 + \gamma^2}, \tag{32}$$

then

$$\bar{\omega}_T = \frac{r_T \omega}{(r_T - 1)\omega + 1} = h(\omega).$$

So

$$V(\bar{\omega}_T \mid \mathbf{x}) \doteq [h'(\hat{\omega})]^2 V(\omega \mid \mathbf{x}), \tag{33}$$

where

$$\hat{\omega} = \frac{\sigma^2}{\sigma^2 + \hat{\gamma}^2},$$

$$h'(\hat{\omega}) = \frac{r_T}{[(r_T - 1)\hat{\omega} + 1]^2}. \tag{34}$$

Equations (31), (33), and (15) imply that the third component of eq. (25) is

$$\frac{r_T^2(\hat{\theta} - I_T)^2}{[(r_T - 1)\hat{\omega} + 1]^4} G. \tag{35}$$

Putting eqs. (26), (30), and (35) together implies that the approximate posterior variance of $\theta_T$ is

$$V_T = (1 - \hat{\omega}_T)(\hat{\theta}_T/e_T) + \hat{\omega}_T^2 \sum_{t=0}^{T} p_t^2[\hat{\gamma}^2 + \hat{\theta}/e_t] + \frac{r_T^2(\hat{\theta} - I_T)^2}{[(r_T - 1)\hat{\omega} + 1]^4} G. \tag{36}$$

### 4.5 QMP algorithm

Here we summarize the QMP formulas. On the right side of the formulas are the section numbers or equation numbers where the formulas were derived.

The audit data for $t = 1, \cdots, T$ is the following:

$Q_t$ = Attribute quality measure in the sample, period $t$ (total defects, defectives, or demerits),
$E_{St}$ = expected value of $Q_t$ given standard quality,
$V_{St}$ = Sampling variance of $Q_t$ given standard quality.

For each period compute the following:

Equivalent defects:

$$x_t = \frac{Q_t}{V_{St}/E_{St}},$$

(Section 3.2)

Equivalent expectancy:

$$e_t = E_{St}^2/V_{St}.$$

For the "prior data" ($t = 0$), let $x_0 = e_0 = 1$.
For $t = 0, \cdots, T$, compute the following: (Section 4.3.2)

Sample index:

$$I_t = x_t/e_t,$$

Weighting factors for computing process average and variance:

$$f_t = \frac{e_t}{1 + e_t/4}, \tag{5}$$

$$g_t = \frac{e_t^2}{2.5 + 1.5e_t + (0.22)e_t^2}, \tag{6}$$

Corresponding weights:

$$p_t = f_t/\sum f_t, \tag{5}$$

$$q_t = g_t/\sum g_t. \tag{6}$$

Over all periods $t = 0, \cdots, T$ compute the following:

Process average:

$$\hat{\theta} = (\sum p_t I_t) \tag{7}$$

Degrees of freedom:

$$df = \frac{2\,[\sum q_t(1/e_t)]^2}{\sum q_t^2(1/e_t^3 + 2/e_t^2)} - 1, \tag{13}$$

Total observed variance:

$$S^2 = \frac{(14.4)\sigma^2 + (\text{df} + 1)\sum q_t(I_t - \hat{\theta})^2}{9 + \text{df}},$$

(15), (19), (20)

Estimated average sampling variance:

$$\sigma^2 = \sum q_t(I_t/e_t), \tag{8}$$

Variance ratio:

$$R = S^2/\sigma^2, \tag{16}$$

*F* and *G*:

$$a = 4.5 + \frac{\text{df}}{2}, \qquad (14), (19)$$

$$B = \sum_{i=0}^{\infty} T(i), \qquad T(0) = 1,$$

$$T(i) = T(i-1)\left[\frac{aR}{a+i}\right],$$

$$F = \frac{B}{B-1}, \qquad (17), (50)$$

$$G = \frac{1}{RF}\left[\left(\frac{a+1}{aR}\right) - (F-1) - \frac{1}{RF}\right], \qquad (18), (54)$$

Current sampling variance:

$$\sigma_T^2 = \hat{\theta}/e_T \tag{23}$$

Sampling variance ratio:

$$r_T = \sigma_T^2/\sigma^2 \tag{32}$$

Process variance:

$$\hat{\gamma}^2 = FS^2 - \sigma^2 = (FR-1)\sigma^2, \tag{21}$$

$$\hat{\omega}_T = \sigma_T^2/(\sigma_T^2 + \hat{\gamma}^2), \tag{22}$$

Weights:

$$\hat{\omega} = \sigma^2/(\sigma^2 + \hat{\gamma}^2) = 1/FR, \tag{34}$$

Best measure of current quality:

$$\hat{\theta}_T = \hat{\omega}_T\hat{\theta} + (1 - \hat{\omega}_T)I_T, \tag{24}$$

Posterior variance of current quality:

$$V_T = (1 - \hat{\omega}_T)(\hat{\theta}_T/e_T) + \hat{\omega}_T^2 \sum p_t^2\left[\hat{\gamma}^2 + \frac{\hat{\theta}}{e_t}\right] + \frac{r_T^2(\hat{\theta} - I_T)^2}{[(r_T - 1)\hat{\omega} + 1]^4}G, \tag{36}$$

Box chart percentiles:

$$\alpha = \hat{\theta}_T^2 / V_T, \qquad \text{(Section 3.3.3)}$$

$$\tau = V_t / \hat{\theta}_T,$$

$I99\%$, $I95\%$, $I05\%$, $I01\%$ defined by:

$$1 - G_\alpha(I99\%/\tau) = 0.99,$$

$$\vdots \qquad \text{(Section 3.4.1)}$$

$$1 - G_\alpha(I01\%/\tau) = 0.01.$$

## V. QMP DYNAMICS

The Best Measure and the box chart percentiles are nonlinear functions of all the data, so the dynamic behavior of these results can appear to be complex. But this complex behavior is desirable and can be explained. This section characterizes the fundamental dynamics of QMP by example.

### *5.1 Dynamics of sudden change*

Since QMP is partially based on a long run average, it is natural to be concerned about responsiveness of the box chart to sudden change. If there is a sudden degradation of quality, Quality Assurance would like to detect it. If the producer solves a chronic quality problem, they would like their exceptions to disappear. Figures 7 and 8 illustrate the QMP dynamics of sudden change.

The history data in Fig. 7 is a typical history for a product that is meeting the quality standard. The equivalent expectancy of five is average for a manufacturing audit. The history is plotted on a $T$-rate chart along with six possible values for the current $T$-rate (labeled $A$ through $F$). So the current period is anywhere from standard ($T$-rate $= 0$) to well below standard (Index $= 3.24$, $T$-rate $= -5$).

The right side of Fig. 7 shows the six possible current results plotted in QMP box-chart form. The box chart labeled A is the result of combining current result $A$ with the past five periods. The box chart labeled $F$ is the result of combining current result $F$ with the same past history.

As you can see, the QMP result becomes ALERT at about $T$-rate $= -3$ (letter D) and becomes BN at about $T$-rate $= -4$ (letter E). For the $T$-rate method of rating, you would have a BN at $T$-rate $= -3$. The good past history has the effect of tempering the result of a $T$-rate $= -3$.

It is informative to study the relative behavior of the current sample

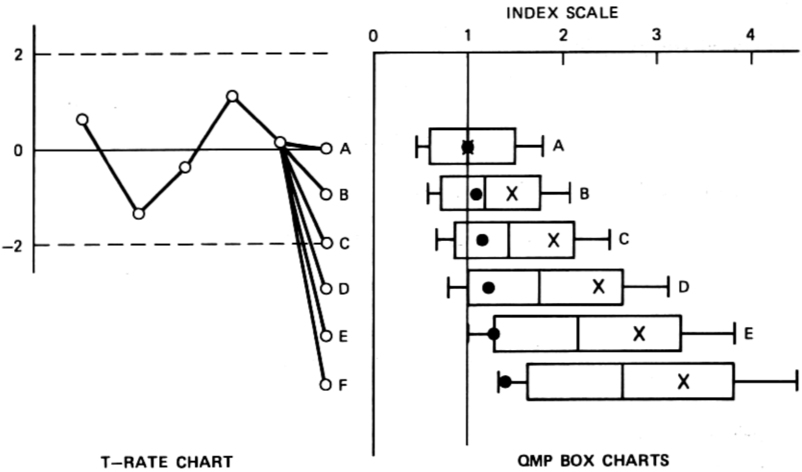


Fig. 7—Dynamics of sudden degradation. The six QMP box charts (labeled $A$ through $F$) result from the analysis of six time series of data, which all have the same *past* history, but have different current values, as shown in the $T$-rate chart. A QMP ALERT is triggered at a $T$-rate of $-3$ (letter D) and a QMP Below Normal is triggered at a $T$-rate of $-4$ (letter E). So a good past history tempers an observed change. Notice that from $A$ to $F$, the Best Measure swings towards the sample value. This results from increasing evidence of an unstable process (expected number of defects equals 5 for this chart).

index, process average, and Best Measure as the current value goes from $A$ to $F$. The current index changes a lot (from 1.00 to 3.24) and the process average changes a little (from 1.00 to 1.38), both in a linear way. The Best Measure also changes substantially, but in a nonlinear way. It changes slowly at first and then speeds up. This is because the weight is changing from 0.71 to 0.32. The weight changes, because as the data becomes more and more inconsistent with the past the process becomes more and more unstable, while the current sampling variance changes slowly in proportion to the process average.

Figure 8 is the dual of Fig. 7. It illustrates the dynamics of sudden improvement. For the first five periods plotted, the process average is centered on an index of two. Then an improvement takes place and from the sixth period on, the sample index is at the standard value of one.

For the first five periods plotted, the rate is BN four times and ALERT once. In the sixth period there is a sudden improvement and the sample index goes to standard. Immediately, there is a jump in the Best Measure and the rate is no longer BN. Because of the increase in process variance, the weight changes from 0.69 to 0.61, putting more weight on the current good result. The posterior variance stays about the same [$\hat{\theta}_T$ gets smaller but $(1 - \hat{\omega}_T)$ gets larger].

For the next five periods the sample index stays at standard. During these periods both the process average and the Best Measure gradually move up towards the standard.

### 5.2 Bogie charts

A Bogie chart is a graphical device for tracking quality assurance audit data during a rating period. Figures 9 and 10 are examples of Bogie charts. The vertical axis is an index scale and the horizontal axis is an equivalent expectancy scale. During the rating period, as the audit sample size builds up, the sample equivalent expectancy increases. So the horizontal axis can also be viewed as a time axis.

The Bogie curves labeled ALERT and BN are plots of the indices in the current sample for which *I*95% and *I*99% (the 95th and 99th percentiles) are exactly one, respectively. So the Bogie curves depend on the past history. The past histories associated with Figs. 9 and 10 have average indices of 0.92 and 4.89, respectively. The variance of the past histories were 0.69 and 5.36, respectively.

To use the Bogie chart, you plot continuously through the period the sample index as a function of the equivalent expectancy in the sample (see Fig. 9). Anytime this plot falls below the ALERT or BN curve, the rate is ALERT or BN at the plotted equivalent expectancy. Then to bail the rate out, the plotted sample index must get above the Bogie curves before the end of the period. For example, in Fig. 9, if the period had ended at an equivalent expectancy of three, then the rate would be ALERT. If it had ended at an equivalent expectancy of five,

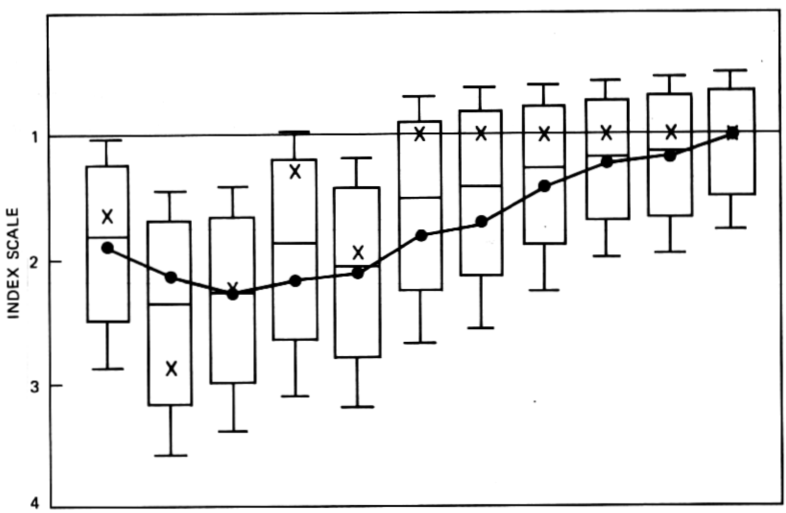


Fig. 8—Dynamics of sudden improvement. As soon as the sample value becomes standard, the product is no longer in the quality exception report (expected number of defects equals 5 for this chart).

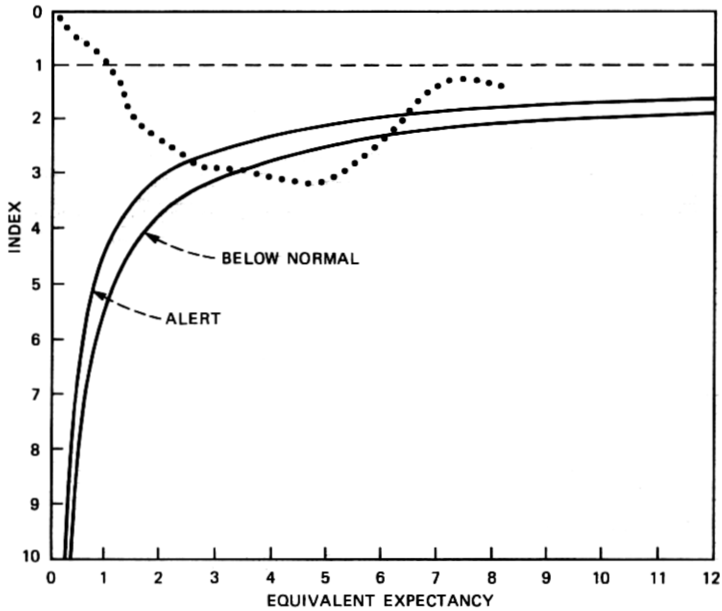


Fig. 9—Index Bogie chart for a good past history. Equivalent expectancy is a measure of how many defects are expected in the sample; so, equivalent expectancy increases with sample size. During a rating period, as the sample size increases, one can track the observed sample index (dotted curve) and compare it to Below Normal and ALERT thresholds.

then the rate would be BN. But the period ended at an equivalent expectancy of eight and there is no exception.

The ALERT Bogie curve in Fig. 10 is interesting. It starts at zero, so you start the period on ALERT. The past history is so bad, that in the

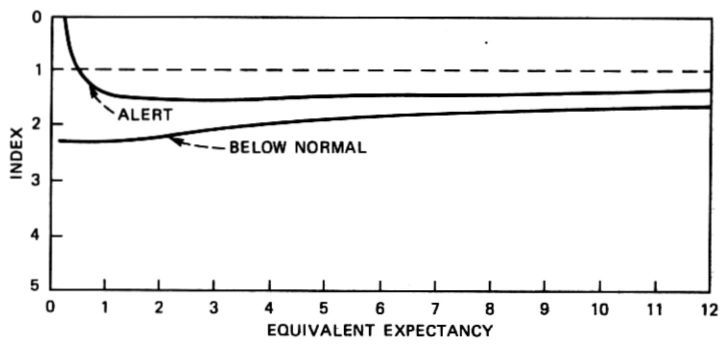


Fig. 10—Index Bogie chart for a substandard past history. The Below Normal and ALERT thresholds are very tight. At the beginning of the period, the product is on ALERT until proven otherwise.

absence of any current data the probability that the current quality will be substandard exceeds 0.95.

### *5.3 Bogie contour plots*

For a fixed past history and current equivalent expectancy, there is a BN Bogie for the current sample index. If the sample index is worse than the BN Bogie, then the product is BN. Figure 11 is a contour plot

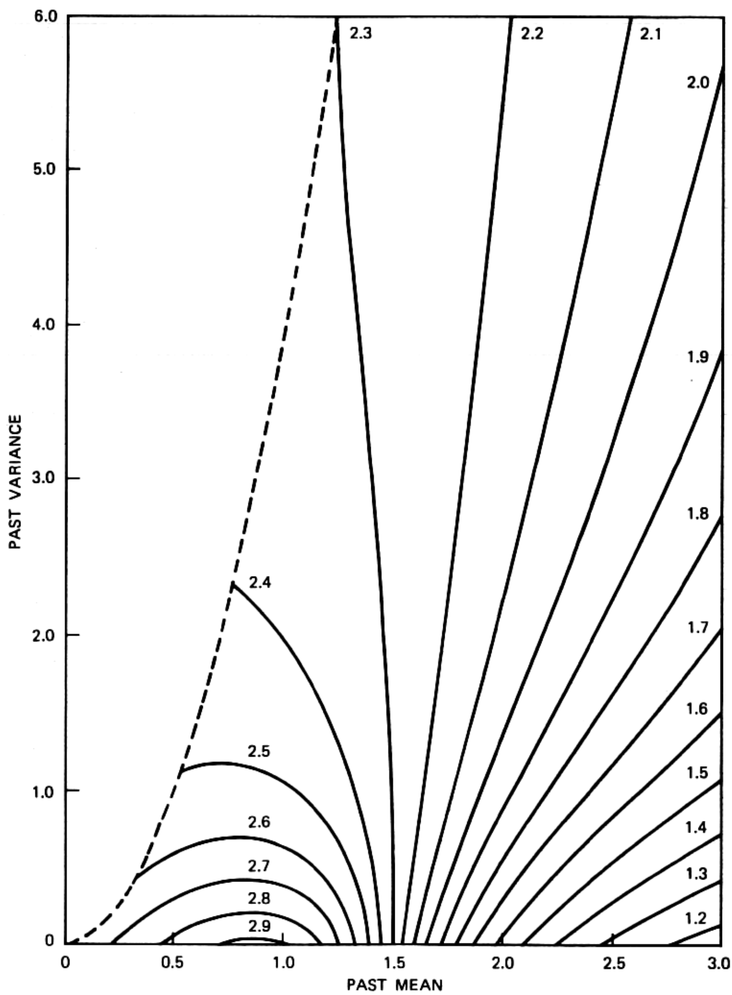


Fig. 11—Below Normal Bogie contour plot. If the past mean is 0.8 and the past variance is 0.7 (on an index scale), then the product is on the contour labeled 2.6. This means that if the current sample index exceeds 2.6, the product will be Below Normal (equivalent expectancy equals 5 for this chart).

of the BN Bogie for an equivalent expectancy of five. The axes are the mean and variance of the five past values of the sample index; i.e.,

$$\bar{I}_p = (1/5) \sum_{t=1}^{5} I_t,$$

$$S_p^2 = (1/5) \sum_{t=1}^{5} (I_t - \bar{I}_p)^2,$$

where $I_t$ is the sample index in past period $t$. For given values of $\bar{I}_p$ and $S_p^2$, we used a standard pattern of $I_t$'s to compute the Bogie. The results are insensitive to pattern. The dashed curve is an upper bound for $S_p^2$.

To see how the contour plot works, consider an example. Suppose $\bar{I}_p = 0.8$ and $S_p^2 = 0.7$. The point (0.8, 0.7) falls on the contour labeled 2.6. This means that if the current sample index exceeds 2.6, then the product will be BN. The contour labeled 2.6 is the set of all pairs ($\bar{I}_p$, $S_p^2$) that yield a BN Bogie of 2.6. The $T$-rate associated with a BN Bogie of 2.6 is −3.6, as shown in Table II.

This contour plot summarizes the BN behavior of QMP for an equiv-

Table II—Index to $T$-rate conversion table

| | $T$-Rate ($T$)* | | |
|---|---|---|---|
| | Equivalent Expectancy ($e$) | | |
| Index ($I$) | 1 | 5 | 10 |
| 1.0 | 0 | 0 | 0 |
| 1.1 | −0.1 | −0.2 | −0.3 |
| 1.2 | −0.2 | −0.4 | −0.6 |
| 1.3 | −0.3 | −0.7 | −0.9 |
| 1.4 | −0.4 | −0.9 | −1.3 |
| 1.5 | −0.5 | −1.1 | −1.6 |
| 1.6 | −0.6 | −1.3 | −1.9 |
| 1.7 | −0.7 | −1.6 | −2.2 |
| 1.8 | −0.8 | −1.8 | −2.5 |
| 1.9 | −0.9 | −2.0 | −2.8 |
| 2.0 | −1.0 | −2.2 | −3.2 |
| 2.1 | −1.1 | −2.5 | −3.5 |
| 2.2 | −1.2 | −2.7 | −3.8 |
| 2.3 | −1.3 | −2.9 | −4.1 |
| 2.4 | −1.4 | −3.1 | −4.4 |
| 2.5 | −1.5 | −3.4 | |
| 2.6 | −1.6 | −3.6 | |
| 2.7 | −1.7 | −3.8 | |
| 2.8 | −1.8 | −4.0 | |
| 2.9 | −1.9 | −4.2 | |
| 3.0 | −2.0 | −4.5 | |
| 3.1 | −2.1 | | |
| 3.2 | −2.2 | | |
| 3.3 | −2.3 | | |
| 3.4 | −2.4 | | |
| 3.5 | −2.5 | | |

* $T = \sqrt{e}(1 - I)$.

alent expectancy of five. As $\bar{I}_p$ gets larger than one, the BN Bogie gets smaller. If $\bar{I}_p$ exceeds 1.6, then the BN Bogie is smaller than 2.34, which corresponds to a $T$-rate of $-3$. So in $T$-rate terms, BN triggers earlier than a $T$-rate of $-3$.

For $\bar{I}_p$ less than 1.4, as $S_p^2$ gets larger, the BN Bogie gets smaller. This is because large $S_p^2$ implies large process variance which makes an observed deviation more likely to be significant.

For very small $S_p^2$, as you move from $\bar{I}_p = 0$ to $\bar{I}_p = 1$, the BN Bogie increases from 2.6 ($T$-rate = $-$ 3.6) to 2.9 ($T$-rate = $-4.2$). This is an apparent paradox. The better the process average, the less cushion the producer gets.

This is *not* a paradox, but an important characteristic of QMP. Remember with QMP we are making an inference about current quality, not long-run quality. If we have a stable past with $\bar{I}_p = 0.2$, and we suddenly get a sample index of 2.7, then this is very strong evidence that the process has changed and very probably become worse than standard. If we have a stable past with $\bar{I}_p = 1$, and we suddenly get a sample defect index of 2.7, then the evidence of change is not as strong as with $\bar{I}_p = 0.2$. The weight we put on the past data depends on how consistent the past is with the present.

Notice that the maximum BN Bogie is 2.92 and occurs at $\bar{I}_p = 0.85$ and $S_p^2 = 0$. It would be a *mistake* for the producer to conclude from the contour plot that he should control his process at $\bar{I}_p = 0.85$ and $S_p^2 = 0$. He cannot achieve $S_p^2 = 0$. The sample index has substantial sampling variance that the producer cannot control.

The Bogie contour plots provide the engineer with a manual tool to forecast the number of demerits that will be allowed by the end of a period. So we have published a book of BN and ALERT Bogie contour plots for equivalent expectancies from 0.5 to 25.

### 5.4 Nonlinearity of QMP

It is tempting to conjecture that if both the process average and the current sample index for one rate are worse than for another, then the Best Measure will also be worse. This is because the Best Measure is a weighted average between the process average and the current sample index. But, since the weight depends on the data nonlinearly, the conjecture is *not true*.

To illustrate this, consider Fig. 12. The six sample indices in Chart B are uniformly worse than the six sample indices in Chart A. But the Best Measure in Chart B is better than for Chart A. The reason is that the weight in B is 0.54 vs 0.12 for Chart A.

## VI. OPERATING CHARACTERISTICS

The $T$-rate and QMP methods of rating are similar in some respects, but there are major differences. In this section, these differences are

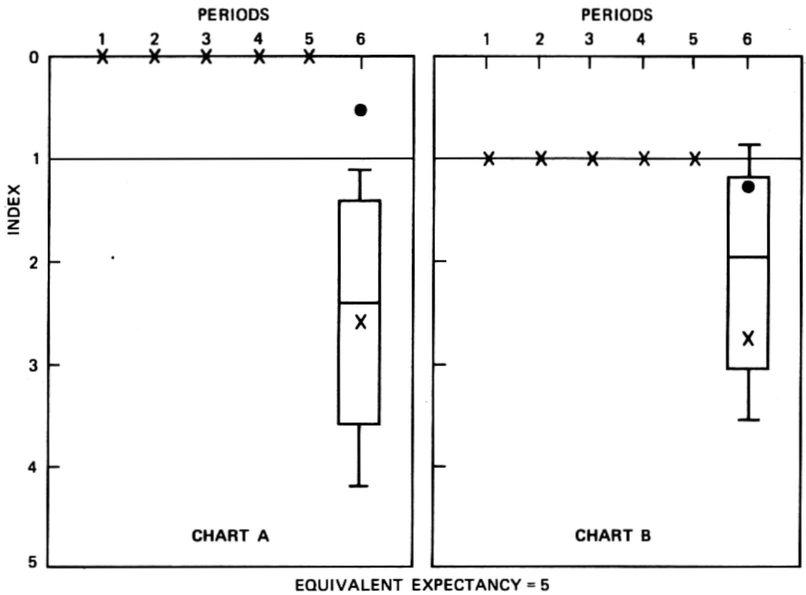


Fig. 12—Nonlinearity of QMP. The sample indices in Chart B are uniformly worse than the sample indices in Chart A; but, the QMP result in Chart B is better than for Chart A. In Chart A, the data provides very strong statistical evidence of an unstable process, so the past data is used very little in estimating current quality. This is not as pronounced in Chart B.

explored using operating characteristics. The differences are a result of different rating formulas and assessment practices.

### *6.1 Ranges of probability substandard for T-rate exceptions*

A QMP analysis of a rating class provides a probability, PS, that the rating class is substandard. For a typical rating period analyzed in detail, we computed PS for all $T$-rate Below Normals and ALERTS. Table III shows the results.

So we find that for $T$-rate BNs, the QMP PS is typically high (greater than 0.97); but, there can be an occasional low PS (e.g., 0.75). However, for $T$-rate ALERT's, the QMP PS is frequently low (e.g., 0.85). This is because the $T$-rate ALERT is an indicator of long-run quality, not current quality.

Table III—Ranges of probability substandard for $T$-rate exceptions

| Exception | Range of Probability | Outlier Probability |
|---|---|---|
| Below Normal | 0.97 → 1.00 | 0.75 |
| ALERT | 0.83 → 0.99 | 0.59 |

### 6.2 Producer's risk and exceptions

Any list of rating classes that is put in an exception report has a producer's risk. It is the fraction of rating classes on the list whose population quality meets the standard. For a given period, let $\theta_i$ = population index, rating class $i$, $i = 1, \cdots, I$. Label the rating classes so that product 1 through product $L$ are on the exception list.

Having done QMP for each rating class, we have a posterior distribution for each $\theta_i$. Now let

$$u_i = \begin{cases} 1, & \text{if } \theta_i \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The number of rating classes on the list whose population quality meets the standard is

$$\sum_{i=1}^{L} u_i,$$

with posterior expected value

$$\sum_{i=1}^{L} \Pr\{\theta_i \leq 1\}.$$

Hence,

$$[\text{producer's risk}]^* = \frac{\sum_{i=1}^{L} \Pr\{\theta_i \leq 1\}}{L}.$$

In QMP, there is an exception list for each threshold probability (TP). TP = 0.95 corresponds to the list of all QMP BNs and ALERTs. Figure 13 shows the QMP producer's risk and number of exceptions as a function of TP for the manufacturing audits in a particular period. The smaller TP, the bigger the exception list and the bigger the producer's risk. Also, note that the producer's risk must be less than 1-TP.

The set of all $T$-rate BNs and ALERTs is another exception list, whose producer's risk is 0.037. This is relatively large because some individual ALERTs have relatively large probabilities (e.g., 0.15) of being standard. The number of $T$-rate exceptions (BN + ALERT) is shown to be 34.

Of course to implement QMP, a particular TP had to be chosen. The TP that would match the $T$-rate producer's risk is about 0.885. But that would lead to an unreasonable (70 percent) increase in exceptions, and a producer's risk of 0.037 is considered too high for this type of exception reporting, because of the high cost of false alarms. So we took TP = 0.95, a reasonable balance between producer's risk and size of the exception report.

* This is not the classical definition of producer's risk.

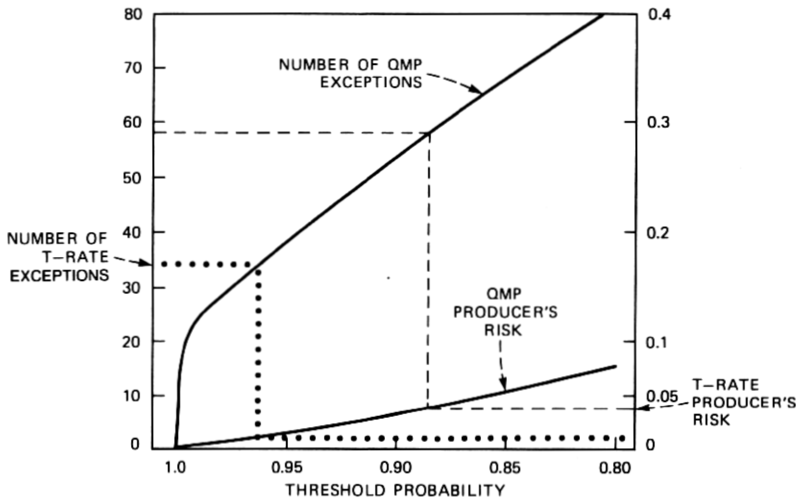


Fig. 13—Operating characteristics of QMP versus the $T$-rate. As the QMP threshold probability for exceptions (currently set at 0.95) is lowered, the number of exceptions and the producer's risk (for a particular rating period) both increase. The number of exceptions and producer's risk for the $T$-rate were 34 and 0.037. For threshold probabilities between 0.96 and 0.89, QMP has more exceptions and lower producer's risk than the $T$-rate.

It should be recognized that these curves depend on the particular set of audits being analyzed. For example, the curves depend on the audit sample sizes. It would be possible to lower sample sizes, decrease the threshold probability, and still maintain a comparably sized exception report with a reasonable producer's risk.

Note that consumer's risk is not analyzed in this paper. Consumer's risk is more relevant to acceptance sampling than to an audit. The main purpose of the audit is to provide quality results to management including a compact exception report of high integrity. The Western Electric quality control organizations have primary responsibility for the quality of each individual lot of product.

## VII. EXAMPLES OF QMP

Here we explore specific examples that illustrate the similarities and differences between QMP and the $T$-rate. In the examples, both QMP and $T$-rate results are based on the same defect data. For the actual implementation of QMP, the defect assessment rules will be different than they are for the $T$-rate as explained in Section 3.1. The intent of this section is to compare how the two rating methods work on the same data.

The examples are shown in Figs. 1, 2, and 14 through 17. These figures show a comparison between the time series of $T$-rates and QMP box charts. Table IV contains summaries of the QMP calculations for the particular periods that will be discussed in the following text.

The QMP calculations shown do not use 1976 data. The box chart for the first period of data available is not shown except in Fig. 15. Period 7706 is the first period for which five periods of past data are used in the QMP box charts. So the comparisons made in this section will involve periods 7706 through 7808.

### *7.1 Agreement with T-rate*

Figure 14 illustrates a $T$-rate borderline* in 7806 preceded by a good history. Since the equivalent expectancy (2.78) is fairly small and the process is fairly stable, the Best Measure (1.81) is heavily weighted (0.65) towards the process average (1.32). The posterior variance (0.36) is fairly large, so I95% is better than standard. However, in the next period, the $T$-rate plummets to −4.8 and the process average drops to 1.77. Now the rate is clearly BN.

### *7.2 Disagreement with T-rate*

In Fig. 15, 7802, the $T$-rate is −3.8 (BN) but there is no exception for QMP. One reason is that QMP is based on the assumption that equivalent defects have a Poisson distribution. A $T$-rate of −3.8 is very significant for a normal distribution, but not as significant for a Poisson distribution with an equivalent expectancy of 0.29. For a normal distribution, the probability, given standard quality, of being below −3.8 is 0.000072. Now the observed number of equivalent defects in 7802 is 2.36. The approximate Poisson probability of exceeding 2.36 equivalent defects given an equivalent expectancy of 0.29 is 0.15—very different from 0.000072.

Another reason is that the QMP result for 7802 is based on one period of data. Rather than using the sample defect index (8.00) as the process average, we use a Bayes estimate [eq. (7)] of 2.77.

Figure 16 is a similar example. In 7708 the $T$-rate of −2.8 is BN because in 7705 the $T$-rate was −2.7. But again, the −2.8 $T$-rate overstates the significance. The equivalent expectancy is only 0.23. Also, the weight (0.60) on the process average (1.81) adjusts the sample index (6.81) to the more moderate Best Measure (3.83). This, together with the large posterior variance (8.47), implies a comfortable $I$95% of 0.56.

Figure 1 illustrates how two similar $T$-rates, both on ALERT, can be either a QMP BN or normal. Compare 7708 with 7804. The sample

* $-3 \leq T$-rate $< -2$, but a good history.

Table IV—Summary data for QMP examples

| Figure | Period | $I_T$ | $\hat{\theta}$ | $e_T$ | $\sigma_T^2$ | $S^2$ | $\sigma^2$ | $\hat{\gamma}^2$ | $\hat{\omega}_T$ | $\hat{\theta}_T$ | $V_T$ | I95% | I99% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 7806 | 2.74 | 1.32 | 2.78 | 0.48 | 0.45 | 0.42 | 0.25 | 0.65 | 1.81 | 0.36 | 0.95 | 0.71 |
| 14 | 7807 | 3.52 | 1.77 | 3.60 | 0.49 | 0.89 | 0.55 | 0.50 | 0.50 | 2.65 | 0.56 | 1.55 | 1.22 |
| 15 | 7802 | 8.00 | 2.77 | 0.29 | 9.55 | 6.80 | 3.85 | 3.55 | 0.73 | 4.19 | 6.97 | 0.47 | 0.96 |
| 16 | 7708 | 6.81 | 1.81 | 0.23 | 7.81 | 9.64 | 7.44 | 5.30 | 0.60 | 3.83 | 8.47 | 0.56 | 0.21 |
| 1 | 7708 | 1.57 | 2.00 | 7.19 | 0.29 | 0.47 | 0.24 | 0.26 | 0.51 | 1.79 | 0.15 | 1.20 | 1.009 |
| 1 | 7802 | 1.61 | 1.51 | 7.19 | 0.21 | 0.44 | 0.21 | 0.26 | 0.45 | 1.57 | 0.14 | 1.01 | 0.84 |
| 1 | 7804 | 1.50 | 1.32 | 6.71 | 0.20 | 0.16 | 0.19 | 0.10 | 0.67 | 1.38 | 0.091 | 0.93 | 0.78 |
| 2 | 7707 | 1.04 | 1.35 | 4.92 | 0.28 | 0.14 | 0.29 | 0.16 | 0.70 | 1.26 | 0.11 | 0.76 | 0.61 |
| 17 | 7807 | 3.10 | 1.10 | 1.40 | 0.78 | 1.36 | 0.91 | 0.75 | 0.51 | 2.08 | 1.02 | 0.74 | 0.46 |

indices of 1.57 and 1.50 are very similar, but the process averages of 2.00 and 1.32 are very different and the weights of 0.51 and 0.67 are different. Hence, the Best Measures are very different and the conclusions are very different.

Figure 2 illustrates a "weak" ALERT under the $T$-rate. The $T$-rate in 7705 through 7707 are −0.1, −0.2, and −0.1, respectively. Although it

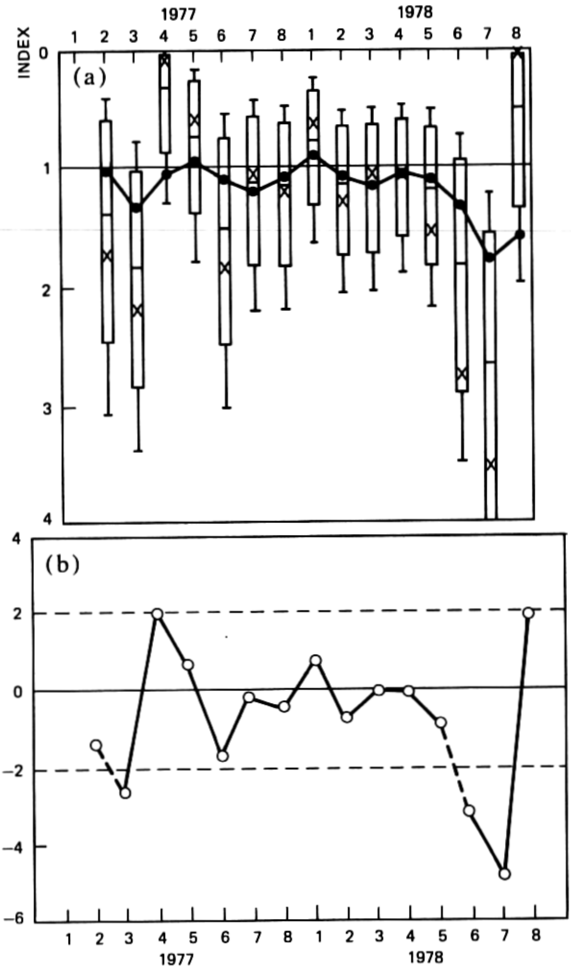


Fig. 14—Example of agreement. Throughout 1978, QMP and the $T$-rate are in agreement. The drop in the sixth period was called "borderline" under the $T$-rate, because it was the first excursion below −2 and it was moderate. The QMP box chart conveys the same borderline message. In the seventh period, the product was Below Normal for both systems.

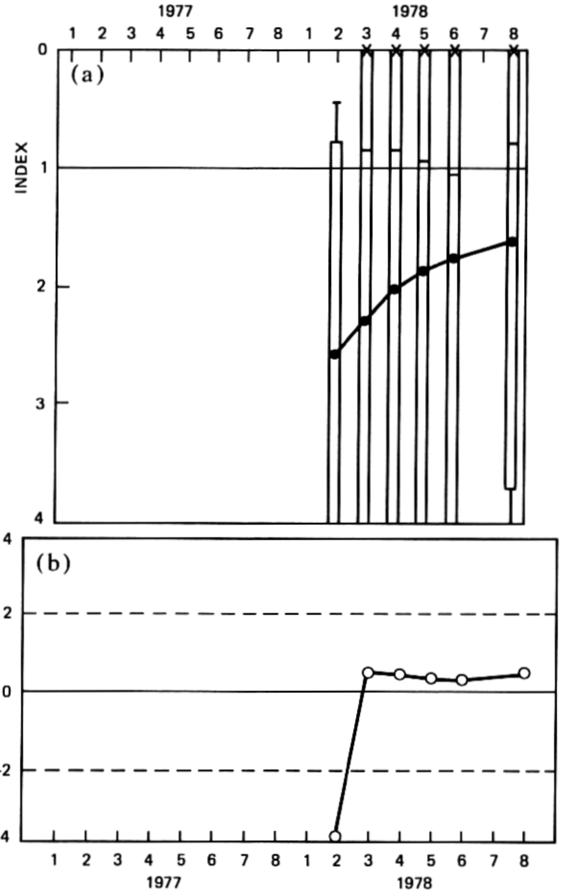


Fig. 15—Poisson versus Gaussian assumption. In the second period of 1978, the expected number of defects in the sample was 0.29 and the observed number of equivalent defects was 2.36. Under the Gaussian assumption, the observed significance level is 0.000072 (i.e., $T$-rate = −3.8); but, under Poisson, the level is 0.15. This explains why the QMP box chart contains the standard.

is unlikely that the quality standard was being met in *every* period from 7702 through 7707, it is *not* unlikely (probability of 0.23) that the quality standard was being met in 7707.

### 7.3 Modification treatment

The $T$-rate system had modification treatments that resulted from the statistical deficiencies of the $T$-rate (see Section 2.8). There are no modification treatments in QMP. The Poisson model and the stabilizing effect of shifting the sample index towards the process average alleviate the need for modification treatments.

In Fig. 17, the 7807 unmodified *T*-rate is −2.5. It is modified to +0.6 because of the "isolated" A weight (100 demerits) defect. Under QMP, the process average (1.10) is only slightly substandard, the weight (0.51) is medium, and the equivalent expectancy (1.40) is small. All this implies a safe *I*95% (0.74) without modification.

### *7.4 Venn diagram of BNs and ALERTs*

In the Venn diagram of Fig. 18, BNs are shown by circles and ALERTS are shown by rectangles. QMP results are shown by dashed lines and *T*-rate results are shown by solid lines. Every rating class that is BN or

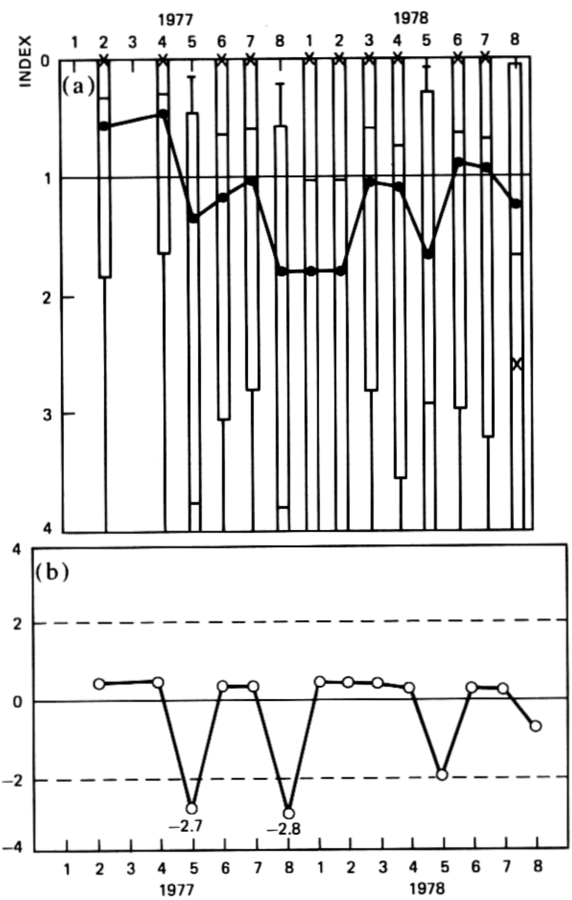


Fig. 16—Statistical jitter in the *T*-rate. With small samples and zero defects, the *T*-rate is slightly larger than zero. Every time a defect is found, the *T*-rate jitters. The message in the QMP chart is that there is too much uncertainty to reach any conclusions.

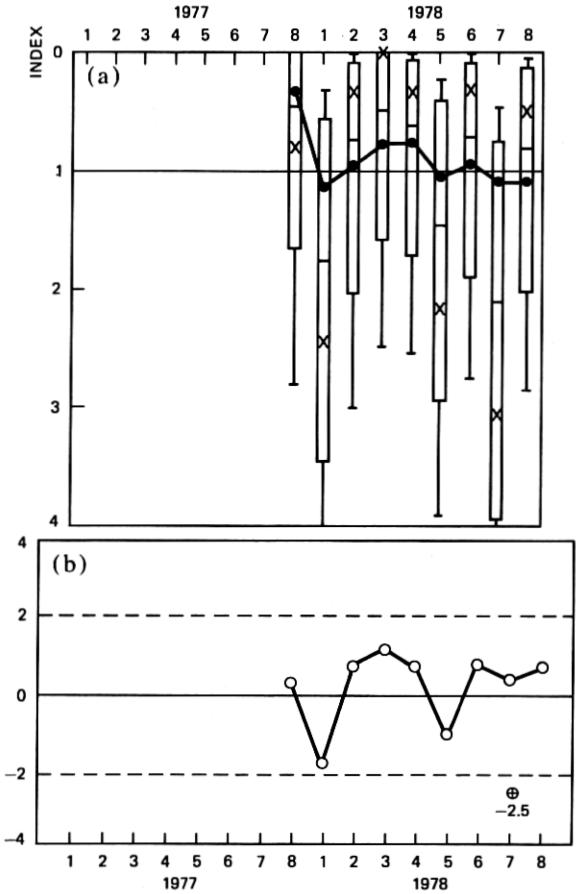


Fig. 17—A case of $T$-rate modification treatment. Because the $T$-rate is biased for small samples, modification treatments were needed to compensate (seventh period, 1978). QMP mathematics obviates the need for modification.

ALERT under QMP or the $T$-rate is represented in the Venn diagram.

Ten rating classes were BN under both methods of rating. Five rating classes were BN under the $T$-rate but ALERT under QMP. There were 16 rating classes that were ALERT under the $T$-rate but normal under QMP. This indicates a major difference. ALERT under the $T$-rate is strong evidence that the quality standards for the current period or some of the past periods have not been met. But it *does not* necessarily imply strong evidence that the quality standard for the current period has not been met. ALERT under QMP implies more than a 95 percent chance that the current quality standards have not been met.

## VIII. ACKNOWLEDGMENTS

Many members of the Bell Laboratories Quality Assurance Center and the Western Electric Quality Assurance Directorate have made important contributions to the development of QMP.

B. T. Howard and E. Fuchs originally gained the support of the senior management of Bell Laboratories and Western Electric. R. A. Peters and L. E. Bray followed this with many presentations to Western Electric management. The role of liason among all the organizations involved was handled by J. M. Wier.

As for technical contributions, R. A. Senior did the original data analyses that transformed the idea into a concrete proposal. C. S. Sherrerd developed the prototype QMP reporting system. S. G. Crawford developed the summary reports and numerical analyses algorithms that permit efficient computation of a thousand QMP charts each period. S. W. Roberts, Jr., developed the QMP defect assessment practices. M. S. Phadke made contributions to the mathematical theory of QMP.

Several members of the Quality Information Systems groups have been developing the data bases and software associated with the official implementation of QMP. They are E. W. Hinds, V. A. Partridge, G. D. Rosen, P. A. Douglas, J. H. Carey, P. D. Ting, and S. Kadakia.

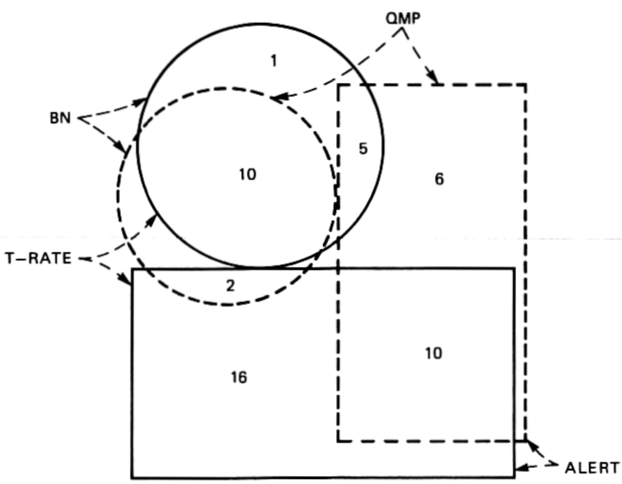


Fig. 18—Venn diagram of exceptions. The Venn diagram accounts for all QMP and *T*-rate exceptions for a particular period using defects assessed under the *T*-rate. The lists of ALERTS under the two systems are quite different (only 10 out of 32 in common).

Western Electric Quality Assurance Headquarters people administered and analyzed the QMP trial and prepared the material for numerous presentations. They are C. Popik, N. O. Dickerson, N. Linardakis, T. M. Ferme, D. Snyder, H. M. Cook, E. Hoffman, and S. Chory.

## APPENDIX A

### *The Gamma Distribution*

A random variable $Y$ has a standard gamma distribution if

$$\Pr\{Y \le y\} = G_\alpha(y) = \int_0^y \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}\, dx,$$

$$\alpha = \text{shape parameter.} \tag{37}$$

A random variable $X = \tau Y$ has a gamma distribution with shape parameter $\alpha$ and scale parameter $\tau$. We write

$$X \sim \text{Gamma}(\alpha, \tau)$$

and

$$\Pr\{X \le x\} = G_\alpha\left(\frac{x}{\tau}\right).$$

The probability density of $X$ is

$$\frac{1}{\tau^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\tau}.$$

The mean and variance of $X$ are

$$E(X) = \tau\alpha, \qquad V(X) = \tau^2\alpha;$$

hence

$$\alpha = E^2(X)/V(X), \qquad \tau = V(X)/E(X).$$

A chi-squared random variable with $\nu$ degrees of freedom has a Gamma distribution; namely,

$$\chi^2_\nu \sim \text{Gamma}\left(\frac{\nu}{2}, 2\right).$$

## APPENDIX B

### *The Poisson-Gamma Bayesian Model*

*Theorem B.1: Assume*

$$x_t \mid \theta_t \sim \text{Poisson}(e_t\theta_t),\ e_t \textit{ known},\ \theta_t \textit{ unknown}$$

*and*

$$\theta_t \sim \text{Gamma}\left(\frac{\theta^2}{\gamma^2}, \frac{\gamma^2}{\theta}\right)$$

$$(i.e., mean = \theta, variance = \gamma^2).$$

*Then*

$$\theta_t \,|\, x_t \sim \text{Gamma}\left(\frac{\bar{\theta}_t^2}{V_t}, \frac{V_t}{\bar{\theta}_t}\right),$$

*where*

$$\begin{aligned}
\bar{\theta}_t &= E(\theta_t \,|\, x_t) \\
&= \omega_t \theta + (1 - \omega_t) I_t, \\
I_t &= x_t / e_t, \\
\omega_t &= \frac{\theta / e_t}{\theta / e_t + \gamma^2}, \\
V_t &= V(\theta_t \,|\, x_t) \\
&= (1 - \omega_t) \bar{\theta}_t / e_t.
\end{aligned}$$

*Proof*: The sampling distribution of $x_t$ is:

$$f(x_t \,|\, \theta_t) = \frac{(e_t \theta_t)^{x_t} \exp\{-e_t \theta_t\}}{x_t!}. \tag{38}$$

The process (prior) distribution of $\theta_t$ is*:

$$p_0(\theta_t) = \frac{e_0{}^{x_0}}{\Gamma(x_0)} \theta_t^{x_0 - 1} e^{-e_0 \theta_t}, \tag{39}$$

$$\theta = x_0 / e_0, \qquad \gamma^2 = x_0 / e_0^2.$$

By Bayes theorem, the posterior density of $\theta_t$ is proportional to the product of equations (39) and (38), which is in turn proportional to

$$[\theta_t^{x_0 - 1} e^{-e_0 \theta_t}][\theta_t^{x_t} e^{-e_t \theta_t}] = \theta_t^{x_0 + x_t - 1} \exp[-(e_0 + e_t)\theta_t]. \tag{40}$$

We recognize eq. (40) as proportional to a Gamma density. So the posterior distribution is Gamma with shape parameter $x_0 + x_t$ and scale parameter $1/(e_0 + e_t)$. And the posterior mean and variance are

$$\bar{\theta}_t = \frac{x_0 + x_t}{e_0 + e_t}, \tag{41}$$

$$V_t = \frac{\bar{\theta}_t}{e_0 + e_t}. \tag{42}$$

---

* Here, $x_0$ and $e_0$ are *not* the same as the "prior data" introduced in Section 4.3.2.

Now multiply the numerator and denominator in both eqs. (41) and (42) by $\theta/e_0e_t$. Theorem B.1 follows. Q.E.D.

## APPENDIX C

### *Chi-Square, Gamma Bayesian Model*

*Theorem C.1: Assume there is a statistic, ss, for which*

$$\frac{\omega}{\sigma^2}(\text{ss}) \Big| \omega \sim \chi^2_\nu \qquad (\textit{chi-square, } \nu \textit{ degrees of freedom})$$

$$\sigma^2 \textit{ known,} \qquad \omega \textit{ unknown}$$

*and*

$$\omega \sim \text{Gamma}\left(a_0, \frac{\sigma^2}{b_0}\right), \qquad a_0, b_0 \textit{ known.}$$

*Then*

$$\omega \,|\, \text{ss} \sim \text{Gamma}\left(a, \frac{\sigma^2}{b}\right),$$

*where*

$$a = a_0 + \frac{\nu}{2},$$

$$b = b_0 + \frac{\text{ss}}{2}.$$

*Proof*: The sampling density of ss given $\omega$ is

$$f(\text{ss} \,|\, \omega) = \frac{1}{(2\sigma^2/\omega)^{\nu/2}\Gamma(\nu/2)}(\text{ss})^{(\nu/2)-1}\exp\left[-\left(\frac{\text{ss}}{2\sigma^2/\omega}\right)\right]. \qquad (43)$$

The prior density of $\omega$ is

$$\rho_0(\omega) = \frac{1}{(\sigma^2/b_0)^{a_0}\Gamma(a_0)}\omega^{a_0-1}\exp\left[-\left(\frac{\omega}{\sigma^2/b_0}\right)\right]. \qquad (44)$$

By Bayes theorem, the posterior density of $\omega$ is proportional to the product of eqs. (43) and (44):

$$\left\{\omega^{\nu/2}\exp\left[-\left(\frac{\omega}{2\sigma^2/\text{ss}}\right)\right]\right\}\left\{\omega^{a_0-1}\exp\left[-\left(\frac{\omega}{\sigma^2/b_0}\right)\right]\right\}$$

$$= \omega^{a-1}\exp\left[-\left(\frac{\omega}{\sigma^2/b}\right)\right].$$

Q.E.D.

*Definition*: Let $X \sim \text{Gamma}(\alpha, \tau)$. Denote the conditional distribution of $X$ given $X \le c$ by

$$C-\text{Gamma}(\alpha, \tau, c).$$

*Corollary*: If instead,

$$\omega \sim C-\text{Gamma}\left(a_0, \frac{\sigma^2}{b_0}, 1\right),$$

then

$$\omega \mid \text{ss} \sim C-\text{Gamma}\left(a, \frac{\sigma^2}{b}, 1\right).$$

*Theorem C.2*: *If* $\omega \sim C-\text{Gamma}(a, \sigma^2/b, 1)$, *then*

$$E(\omega) = \frac{1}{RF},$$

$$V(\omega) = G,$$

*where*

$$R = \frac{b/a}{\sigma^2},$$

$$F = \frac{G_a(aR)}{G_{a+1}(aR)} \qquad \text{[see (37)]},$$

$$G = \frac{1}{RF}\left[\left(\frac{a+1}{a}\right)\frac{G_{a+2}(aR)}{RG_{a+1}(aR)} - \frac{1}{RF}\right]. \tag{45}$$

*Proof*: Note

$$\frac{b}{\sigma^2}\omega = aR\omega \sim C-\text{Gamma}\left(a, 1, \frac{1}{aR}\right).$$

So

$$\begin{aligned}
E(\omega) &= \frac{1}{aR}E(aR\omega) \\
&= \frac{1}{aRG_a(aR)}\int_0^{aR} y\frac{y^{a-1}e^{-y}}{\Gamma(a)}\,dy \\
&= \frac{\Gamma(a+1)}{aRG_a(aR)\Gamma(a)}\int_0^{aR}\frac{y^{(a+1)-1}e^{-y}}{\Gamma(a+1)}\,dy \\
&= \frac{aG_{a+1}(aR)}{aRG_a(aR)} \\
&= 1/RF.
\end{aligned} \tag{46}$$

Now

$$
\begin{aligned}
E(\omega^2) &= \frac{1}{(aR)^2} E[(aR\omega)^2] \\
&= \frac{1}{(aR)^2 G_a(aR)} \int_0^{aR} y^2 \frac{y^{a-1}e^{-y}}{\Gamma(a)}\, dy \\
&= \frac{\Gamma(a+2)}{(aR)^2 G_a(aR)\Gamma(a)} \int_0^{aR} \frac{y^{(a+2)-1}e^{-y}}{\Gamma(a+2)}\, dy \\
&= \frac{(a+1)aG_{a+2}(aR)}{(aR)^2 G_a(aR)} \\
&= \frac{(a+1)G_{a+1}(aR)G_{a+2}(aR)}{aR^2 G_a(aR)G_{a+1}(aR)} \\
&= \left(\frac{a+1}{a}\right) \frac{G_{a+2}(aR)}{R^2 F G_{a+1}(aR)}.
\end{aligned}
\tag{47}
$$

This along with eq. (46) implies $V(\omega) = G$.

***Computational formula for F***

Let

$$g_a(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}\, dx.$$

From Ref. 11, page 262, 6.5.21,

$$G_{a+1}(x) = G_a(x) - \left(\frac{x}{a}\right) g_a(x). \tag{48}$$

Now define

$$B_a(x) = \sum_{i=0}^{\infty} T(i), \qquad T(o) = 1,$$

$$T(i) = T(i-1)\left[\frac{x}{a+i}\right] = 1 + \frac{x}{a+1} + \frac{x^2}{(a+1)(a+2)} + \cdots.$$

By Ref. 12, page 3,

$$G_a(x) = \left(\frac{x}{a}\right) g_a(x) B_a(x). \tag{49}$$

Putting eqs. (48) and (49) together implies

$$F_a(x) \triangleq \frac{G_a(x)}{G_{a+1}(x)}$$

$$= \frac{(x/a)g_a(x)B_a(x)}{(x/a)g_a(x)B_a(x) - (x/a)g_a(x)}$$

$$= \frac{B_a(x)}{B_a(x) - 1}.$$

So

$$F = \frac{B_a(aR)}{B_a(aR) - 1}. \tag{50}$$

***Computational formula for G***

Directly by definition, it follows that

$$B_{a+1}(x) = \left(\frac{a+1}{x}\right)[B_a(x) - 1]. \tag{51}$$

Therefore,

$$\frac{1}{F_{a+1}(x)} = 1 - \frac{1}{B_{a+1}(x)}$$

$$= 1 - \frac{1}{[(a+1)/x][B_a(x) - 1]}$$

$$= 1 - \left(\frac{x}{a+1}\right)[F_a(x) - 1]. \tag{52}$$

Now plug eq. (52) into the first term in the square bracket of eq. (45) and get

$$\left(\frac{a+1}{a}\right)\frac{G_{a+2}(aR)}{RG_{a+1}(aR)} = \left(\frac{a+1}{a}\right)\frac{1}{RF_{a+1}(aR)}$$

$$= \left(\frac{a+1}{a}\right)\frac{1}{R}\left[1 - \left(\frac{aR}{a+1}\right)(F - 1)\right]$$

$$= \frac{a+1}{aR} - (F - 1). \tag{53}$$

So

$$G = \frac{1}{RF}\left[\left(\frac{a+1}{aR}\right) - (F - 1) - \frac{1}{RF}\right]. \tag{54}$$

## APPENDIX D

***Moments of functions of the sample index***

If

$$x_t \mid \theta_t \sim \text{Poisson}(e_t\theta_t),$$

$$\theta_t \mid \theta, \gamma^2 \sim \text{Gamma}\left(\frac{\theta^2}{\gamma^2}, \frac{\gamma^2}{\theta}\right),$$

then $x_t \mid \theta, \gamma^2$ is a negative binomial with density

$$f(x_t \mid \theta, \gamma^2) = \frac{\Gamma(x_t + \theta^2/\gamma^2)}{x_t!\Gamma(\theta^2/\gamma^2)} \left[\frac{1}{1 + \theta/e_t\gamma^2}\right]^{x_t} \left[\frac{\theta/e_t\gamma^2}{1 + \theta/e_t\gamma^2}\right]^{\theta^2/\gamma^2}.$$

Let

$$\mu_1 = \text{mean of } x_t,$$

$$\mu_\nu = \nu\text{th central moment of } x_t, \qquad \nu = 2, 3, \cdots.$$

Then according to (Ref. 11, page 929),

$$\mu_1 = \alpha P,$$

$$\mu_2 = \alpha PQ,$$

$$\mu_3 = \alpha PQ[Q + P],$$

$$\mu_4 = \alpha PQ + A(\alpha PQ)^2, \tag{55}$$

where

$$\alpha = \theta^2/\gamma^2,^*$$

$$P = \gamma^2 e_t/\theta,$$

$$Q = 1 + P,$$

$$A = 3 + 6\gamma^2/\theta^2.$$

Now let

$$\xi_1 = \text{mean of } I_t,$$

$$\xi_\nu = \nu\text{th central moment of } I_t, \qquad \nu = 2, 3, \cdots.$$

It follows from (55) that

$$\xi_1 = \theta,$$

$$\xi_2 = \gamma^2 + \frac{\theta}{e_t},$$

$$\xi_3 = \frac{2\gamma^4}{\theta} + \frac{3\gamma^2}{e_t} + \frac{\theta}{e_t^2},$$

$$\xi_4 = A\gamma^4 + \frac{2A\theta\gamma^2}{e_t} + \frac{A\theta^2 + \gamma^2}{e_t^2} + \frac{\theta}{e_t^3}. \tag{56}$$

* A different $\alpha$ from the one in the main text.

An application of these formulas is

$$V[(I_t - \theta)^2] = \xi_4 - \xi_2^2$$
$$= (A-1)\gamma^4 + \frac{2(A-1)\theta\gamma^2}{e_t} + \frac{[(A-1)\theta^2 + \gamma^2]}{e_t^2} + \frac{\theta}{e_t^3}. \quad (57)$$

Now define

$$Y_t = (I_t - \theta)^2 - I_t/e_t.$$

Further applications of (56) are

$$E(Y_t) = \xi_2 - \xi_1/e_t$$
$$= \gamma^2 \quad (58)$$

and

$$V(Y_t) = E(Y_t^2) - \gamma^4$$
$$= E\left[(I_t - \theta)^2 - \frac{1}{e_t}(I_t - \theta) - \frac{\theta}{e_t}\right]^2 - \gamma^4$$
$$= (A-1)\gamma^4 + \frac{[2(A-1)\theta\gamma^2 - 4\gamma^4/\theta]}{e_t}$$
$$+ \frac{[(A-1)\theta^2 - 4\gamma^2]}{e_t^2}. \quad (59)$$

## REFERENCES


1. E. Fuchs and B. T. Howard, "Quality Assurance Tradition and Change," Bell Laboratories Record, *56*, No. 9 (October 1978), pp. 226–31.
2. R. Peters and I. O. Karraker, "Why You Can Trust Your Telephone," Ind. Res. *17*, No. 11 (November 15, 1975), pp. 47–51.
3. W. A. Shewhart, "Nature and Origin of Standards of Quality," B.S.T.J, *37*, No. 1 (January 1958), pp. 1–22.
4. W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, New York: D. Van Nostrand, 1931.
5. H. F. Dodge, "A Method of Rating Manufactured Product," B.S.T.J., *7* (April 1928), pp. 350–68.
6. H. F. Dodge and M. N. Torrey, "A Check Inspection and Demerit Rating Plan," Ind. Qual. Contr. *13*, No. 1 (July 1956), pp. 1–8.
7. B. Efron and C. Morris, "Stein's Paradox in Statistics," Sci. Am., *236*, No. 5 (May 1977), pp. 119–27.
8. C. Stein, "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, California: University of California Press, 1955, pp. 197–206.
9. Bruce Hoadley, "An Empirical Bayes Approach to Quality Assurance," *33rd Annual Technical Conference Transactions of ASAC*, May 14–16, 1979, pp. 257–63.
10. B. Efron and C. Morris, "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach," J. Am. Statist. Ass., *68*, No. 341 (March 1973), pp. 117–30.
11. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover, 1965.
12. M. B. Wilk, R. Gnanadesikan, and M. J. Huyett, "Probability Plots for the Gamma Distribution," Technometrics, *4*, No. 1 (February 1962), pp. 1–20.